Commentary

THE EDUCATION OF ADULTS

THERE is great truth in the aphorism that 'education is what remains when you have forgotten all that you have ever been taught'. In teaching, action and reaction may or may not be equal but certainly they are opposite and there is no captive audience of adults. It is a simple and instructive experience to talk to children about educating parents but it is dangerous to talk to adults about educating adults. They may feel a suggestion that they need it.

Adult Education was never more needed than it is today. At the most obvious level there is so much to know! We are beginning to explore space, beyond the surface of the globe; we have a measure of understanding of nature and natural phenomena; mankind is beginning to have some control over its destiny; we are living longer, healthier lives; we have more leisure. At the same time there is a widespread social dread and in consequence an apathy. There is talk of 'two cultures'; there is alarm at 'excessive specialization'; there is ignorance, frequently a complacent ignorance, of each others affairs. There have always been two cultures, an expansive and a concentric; today it is perhaps more obvious and more dangerous. Skill has always required concentration and specialization. And the ignorance of superstition is more dangerous than the ignorance of apathy. But education, enlightenment, self-enlightenment is needed to counter an apathetic ignorance which will surely grow, which is growing into an ignorance of superstition.

There are benefits associated with the need and opportunity. Technical communications have never been better. Radio and television are found in every home. In the industrialized countries, at least, the literacy is high. Leisure provides the time and the opportunity. There are also dangers. Education is now a passport to commercial success; this is no bad thing but it can lead to a confused heresy, the heresy that it can—and should—be packaged and purchased as an insurance certificate. The old Workers' Educational Association began as a co-operative among the workers to improve their training and education. The W.E.A. has achieved much. There is a great and expanding future for Adult Education, not in the field of vocational training, but in general education, in helping the public to appreciate the excitement of the world they live in, to give them a taste of intellectual exploration. The great expansion of formal education should be accompanied by an expansion of adult education. It is dangerous not to expand for ignorance leads to political irresponsibility.

Science is not a topic or a subject, it is a spectrum of human activities and there are at least three reasons why it should occupy an important place in educating adults. The first reason is that it is a major human activity and offers some of the most striking triumphs of the human mind. It has revolutionized our lives. There is a misconception that science is concerned with gadgets because indeed many material benefits follow its exploitation. The intellectual discipline may be the fundamental reason but the most obvious motive for paying attention to the sciences is the simple, natural curiosity of the public. The third count is the sense of adventure. Science is 'knowledge making', it is an exploration, an exploration of nature, an exploration of the human mind; the applications have been stupendous both in scale and in impact.

The problem of educating adults who are free is quite different from that of educating children who are not; that is why the interest and the motives are so important. Consequently the techniques must be quite different from those used in training professionals. There is a world of difference between professional practice and 'knowing about'. The general public require a corrective to the growing emotional superstition; they are disturbed by the possibilities of nuclear power and alarmed at the prospects of a computer ridden society. The picture of 'electronic brains' conjures up for too many a faceless master rather than an inanimate slave. They require a broad appreciation rather than a detailed understanding. The scientist is not a medicine-man. Some of us require to learn this lesson no less than the layman. After all a physicist's ignorance of biology may be as abysmal as that of an historian.

The success of the popular writings of Jeans and Eddington before the war emphasises that the need is there. There are many means of approach for there is no single answer to the question of educating adults. The mass media of radio and television can can be used to break down the dangerous barrier of superstition. Experiments in their use are being made, but not enough. Many are pretentious with an aura of heavy-breathing which discourages the weak and impresses the faint-hearted. The demonstrations, the talks for small groups, the means of developing the experience of each and every client, these too require experiment. And the well-tried methods with a captive audience are irrelevant!

If we do not do it we are in danger of reaching a stage when three 'A' Levels in Science become the passport to join the witch-doctor's union!

First Prize in the Waverley Gold Medal Essay Competition

THE SOLID IMAGE MICROSCOPE

An Instrument giving Information in Depth by Optical Scanning

R. L. GREGORY

University Lecturer in Experimental Psychology, The Psychological Laboratory, Cambridge

In this essay, which won the 1960 *Waverley Gold Medal* Essay Competition, Mr GREGORY describes the historical background, the general principles and the design problems of a new solid image microscope, and discusses its possible applications.

ALL knowledge is based upon information received by the senses, and all control is ultimately dependent upon movements of limbs. Instruments and tools are extensions of senses and limbs—whether telescopes or microscopes, hammers or cars. As science and technology advance, our senses and limbs become inadequate, and so civilization depends upon devices to increase the input information, and the output precision and power of bodies, themselves unchanged since long before civilized life began. Given instruments and tools, man is a species far removed from any other. Rather than de-humanizing man, technology de-apes him.

As instrument design advances we gain information previously denied us. The sense organs are transducers lying between the world and the central nervous system; instruments are transducers between the world and the senses, and they must be suitably matched to both, their design considerations bridging physics and sensory physiology. Most instruments are in fact extensions of the eye, which in man is the most efficient sense organ, the ear running in second place. The eye may be thought of as several largely independent sensory systems in one organ; for it provides information concerning brightness, colour, movement and form in three spatial dimensions. Now when we extend the functions of the eye by the addition of an instrument, some of these channels are generally sacrificed for technical reasons, and so though improving the matching of the eye to some selected features of the world, we may impair its matching to the brain by cutting out some of the normally functional neural channels, and the resulting perception can be misleading. A case in point, and the case with which we are concerned, is loss of stereoscopic vision in the microscope.

None of the usual cues to depth were available to the early microscopists. They were denied stereoscopic vision—which depends upon disparate images upon the two retinas resulting from slightly different view points—and they could not view objects from different positions by moving the head. The first attempt to provide binocular vision with a microscope was surprisingly early, being made by a French Capuchin friar, le Père Cherubin, this predating any experimental work on stereoscopic vision by nearly two hundred years. He wrote, in 1677, ' Some years ago I resolved to effect what I had long before premeditated, to make a microscope to see the smallest objects with the two eyes conjointly; and this project has succeeded even beyond my expectations; with advantages above the single instrument so extraordinary and so surprising, that every intelligent person to whom I have shown the effect has assured me that inquiring philosophers will be highly pleased with the communication'. This was apparently quite forgotten until Wheatstone's invention of the stereoscope, introduced in 1838. Wheatstone asked both Ross and Powel to construct a stereoscopic microscope, but this was not done, the first instrument of modern times being that of an American, J. L. Riddell, built about 1850 and greatly improved by F. H. Wenham in 1853. Wenham realised that a high power objective derives different views from its central and marginal rays, and succeeded in deriving two disparate images from a single objective, by dividing the exit rays and producing appropriate right and left images for the two eyes corresponding to the different view points. This is the basis of all modern high power stereoscopic microscopes.

Binocular stereoscopic microscopes are used a great deal, but suffer from the essential limitation

The Waverley Gold Medal prizewinner, RICHARD LANGTON GREGORY, M.A. (Cantab) *with the solid image microscope. Mr Gregory was born in London, in 1923, educated at King Alfred School, Hampstead, and after five years in the R.A.F., at Downing College, Cambridge where he read moral sciences for two years and psychology for one year. After three years in the Medical Research Council's Applied Psychology Research Unit at Cambridge, he became a demonstrator and subsequently a lecturer at Cambridge under Professor O. L. Zangwill. He holds research grants from the M.R.C. and D.S.I.R. for work on the special senses and for designing new kinds of biological apparatus. In 1956 he won the CIBA Foundation Prize for work relating to age and in 1958 the Craik Prize in Physiological Psychology*

that it is not possible to produce an image of a thick section at high magnification with the structure in focus throughout the depth of the section. In other words, lenses giving large useful magnification must have but a small depth of field. Thus although the appearance of depth can be given in this way, the method can be used only over a very small range in depth in the object. The experimental microscope to be described does, at least in principle, beat this essential limitation in lenses as they are normally employed.

The General Principle of the 'Solid Image' Microscope

The depth of field of high power microscope objectives is extremely small, the resolution in depth being comparable to the resolution in the image plane. What we do is to accept this, and to use this feature of high power lenses to extract information in depth by scanning the thin plane of sharp focus through the specimen.

As we know from the cinema, if a series of pictures is presented in rapid succession, they fuse. It might thus be thought that if a series of pictures were obtained corresponding to successive layers and presented in rapid succession, we would see the entire three dimensional structure, and that this would be the simple answer to the problem of the limited depth of focus of objectives; but this is not so, for we should see only a muddle of pictures lying upon each other like a series of photographs (though related photographs) taken upon a single film. This then is not the full answer. But now imagine the series of pictures as photographic transparencies placed in order, one behind the next, to form a cube of pictures through which we could look. Would we not have a solid, if layered, photograph in three dimensions? And could we not walk round it and see it from various positions? Would not the structure change by parallax as we change our view point, like a real object lying in space? If this would work, we should have put to advantage the limited depth of focus of microscope objectives, by using the thin plane of sharp focus to obtain separate pictures of the structure at various depths, to give a single 'solid' picture occupying three-dimensional space. Thick specimens could be used, for since the plane of focus moves through the specimen it can, and indeed must, be many times the depth of the plane of focus of the objective. The thinner the plane of focus the better the resolution in depth, and so the original optical limitation has been put to use.

It is in practice a nuisance to have to take a series of photographs, and it is difficult to arrange them so that there are no serious errors of registration between the planes; it would thus be better to provide a solid picture immediately, without the photographic intermediary.

Suppose we project the image obtained by the microscope's objective upon a screen, so that we may see it just as we look at a slide projected on a screen. Suppose now that we move the objective lens up and down, not slowly but fast, the focal plane running regularly up and down through the specimen many times a second. The result would be a blurred muddle—just as a set of photographs lying upon each other in two dimensions would be a muddle—but if, somehow, we could present the image corresponding to each depth at its correct distance from the observer, then we should obtain an image without confusion, and it would be a three-dimensional 'solid' image. This may be done simply by vibrating the screen back and forth, in synchrony with the scan of the focal plane through the specimen. As the screen moves away from the observer, the picture upon it changes to correspond with each new structure revealed in the depth of the specimen by the scan of the plane of focus. The observer may view the solid image lying in the volume swept by the vibrating screen from any frontal position, when he will see the structure change by parallax as though it were a real object: he will see a magnified model of the object lying in a luminous block of space.

We have built just such an instrument, and with it we can obtain 'solid' pictures which do show structure lying in depth. The rough surface of a slide looks like a miniature mountain range; the organs inside an insect may be seen in their relative positions. In place of a fly entombed in amber we have a fly in a block of light, and our fly need not be dead.

Design Problems

Optical Considerations

It is of primary importance for acceptable resolution in depth that the plane of focus of the objective lens be as thin as possible for each point in the object to be represented in its correct position in depth in the image space. It is thus of first importance to examine the theoretical resolution obtainable in depth with microscope objectives.

The depth of field decreases as the aperture of the lens increases—which is fortunate for we also need high aperture to obtain good resolution in the normal image plane if high magnification is to be used, as was established theoretically by Abbe at the end of the nineteenth century. The depth of field, *df*, is determined by the allowed difference in path length from the centre and the periphery of the lens as the distance of the object is changed. Some criterion is needed to assign a value to the allowable difference, and Rayleigh's criterion of one quarter of a wavelength of light is generally accepted; *df* is then given by

$$df = \frac{\text{Allowed difference in path length } (\lambda/4)}{n.\ \sin^2 U/2}$$

where U is the angle made by the marginal rays with the axis in the object space, and n is the refractive index of the medium on the object side of the lens.

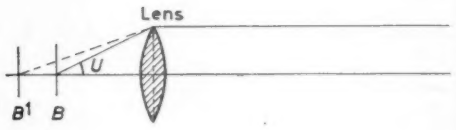

Figure 1. The basic geometry of the depth of field in lenses, showing the changing angle, U, and the central and peripheral rays

Taking the wavelength of light as 0·0005 mm the depth of field of lenses of various apertures may be calculated.

Table 1. Calculated figures giving depth of field for various values of numerical aperture

	Numerical aperture	*Depth of field (mm)*
Working in air	0·25	0·0079
	0·50	0·0019
	0·75	0·0008
Oil immersion	1·00	0·0007
	1·25	0·0004

It will be seen that the depth of field is very small for high apertures, and smaller working in air than oil immersed for equivalent apertures. Thus resolution in depth (which we may term the *Z*-axis) can be quite high—but we must make a correction if we are not to be misled into undue optimism. Rayleigh's criterion is generally taken for just acceptable loss of resolution, but we should demand more than this, for we wish to know when a point will be *invisible*, rather than just detectably fuzzy. For this reason the effective depth of field should probably be increased in each case by a factor of about ten, and we should thus expect resolution in depth to be considerably worse than resolution in the normal image plane. In practice, we should not require resolution in the *Z*-axis of the same order as the other axes, for it is only necessary to locate the position on the *Z*-axis of structures identified in the normal plane, and so this inevitable loss in the *Z*-axis

right be acceptable. It does mean, however, that we can hardly expect image quality to be maintained for angles of view far round the side of the image.

In order to ensure that the objective is being used at its fullest aperture, it is essential to provide it with light from a high aperture sub-stage condenser. In practice the adjustment of this condenser is critical, but no special optics are required.

The choice of objective lens is difficult, for we need not only a large aperture but also a large working distance in order to scan the focal plane deep down into the specimen without coming up against the front surface of the lens. Now in practice high aperture lenses have very small working distances, and this is a major difficulty. Objectives having very long working distances have been made, but not with apertures greater than about N.A. 0·65, and this is not high enough to give acceptable resolution in the *Z*-axis. It is essential to obtain a specially designed lens for our final solid image microscope.

It is necessary to accomplish the following operations, and all with high precision: (*1*) the focal plane must be scanned up and down through the specimen object, at a rate of at least 30 scans per second; (*2*) the screen must be vibrated at the same rate and in phase with the object scan, but with considerably greater amplitude. In fact the amplitude should be greater than the object scan by the magnification used.

We have tried nearly a dozen ways of obtaining the above mechanical movements. We have used tuned forks, a reciprocating engine—the screen being mounted on the piston—and various kinds of non-tuned vibrating elements, including loudspeaker cones. We have used only sinusoidal scans, the mechanical problem of producing precise linear saw-tooth movements of the objective at the rate required being probably greater than is justified by any advantages over sinusoidal movement.

Experiment shows that tuned systems are not satisfactory, though it has taken a year for this conclusion to become clear. They have the advantage that relatively large amplitudes can be obtained with low power—we were able to get a screen amplitude of nearly 1-inch with an expenditure in power of about 10 watts, while the same amplitude in a non-tuned system required 200–300 watts—and this is a serious consideration since the phase has to be controlled, which is difficult and expensive at high powers. Tuned mechanical systems drift in phase with slight temperature changes, and we never succeeded in making a stable arrangement of this kind.

It is not too difficult to impart precise sinusoidal movement of the small amplitude required (a few thousandths of an inch) to the objective lens. The best way we have found so far is to use standard moving coil vibrator units, manufactured for testing components against sustained vibration and metal fatigue. We drive a pair of these from an amplifier fed by a 50 cycles variable phase signal, the lens being mounted in the centre of a beam which vibrates up and down. The screen is more difficult in that the amplitude must be much greater, though the required precision is rather less. To use a non-tuned vibrator unit here is prohibitively expensive, and tuned systems we have ruled out as being too susceptible to spontaneous shifts in phase. We therefore need some new way to obtain the image scan.

It is possible to avoid vibrating the screen physically. This may be done by using a rotating disc of suitable design, and projecting the image on a sector of the disc which changes its position with respect to the observer as the disc rotates.

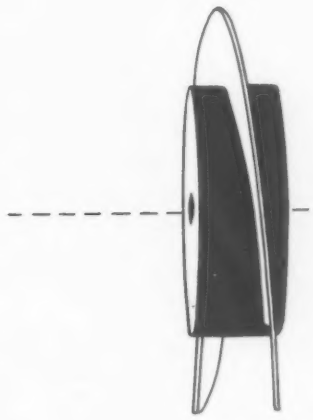

Figure 2. Diagram of a spiral screen. It may be seen how the position of a given sector changes as the screen as a whole rotates round its axis. (GREGORY[4])

The simplest arrangement is to use a flat disc rotated on a shaft fixed to its centre but not quite normal to the plane of the disc. This is a 'swash plate', and an image projected on a sector will move backwards and forwards as the disc revolves, but it

will also change in shape, appearing as a solid wedge. (In practice half the disc is painted black to give a scan only in one direction, and for the other systems the light is chopped off on the return scan to avoid mis-registration of the images derived from the up and down scans, which are generally not quite identical in form. In addition, for sinusoidal scans it is essential to modulate the light source in order to increase the brightness as the velocity increases, otherwise all that is seen is a sandwich consisting of two pictures back and front, with nothing much in between where the screen velocity is high. This is done by using a high pressure mercury arc fed at 50 cycles.)

An arrangement preferable to the flat but tilted disc, though far more difficult to make to the required precision, is a plane helix. This has the great advantage that the picture is not distorted, and this represents the best means of producing the object scan we have so far devised, the helical disc being driven by a synchronous motor locked to the mains and rotating at 3000 r.p.m., half the helix being blacked out, to give 50 one-way scans per second.

A rotating screen of this kind can give a linear scan, though it is difficult to impart this kind of movement to the objective. A reasonable arrangement seems to be to move the objective sinusoidally while the screen gives a linear scan, but to chop the light at the ends of the scan and use the middle only, which is substantially linear. The object and image scans can be brought into synchrony simply by rotating the body of the motor driving the screen until they are in phase. We thus have a very simple arrangement in which amplifiers and phase shift circuits are avoided entirely, and it is stable and free from mechanical troubles.

Finally, it is almost certainly possible, though we have not yet done it, to avoid the screen altogether and give solid pictures by scanning, but viewed directly with a stereoscopic pair of eye pieces. This could be done by providing the object scan as described, but in place of a vibrating or rotating screen giving a real image, the virtual images provided by the eye piece lenses would be oscillated slightly from side to side in phase with the object scan. This would be optically equivalent to viewing the image on a vibrating screen, and would have the advantage that images sufficiently bright could be obtained with dark ground or phase contrast lighting, which is not the case with projection upon a screen. We might also expect an improvement in image quality, though it would not be a simple matter to view the image from various positions.

Applications of the Solid Image Microscope to Science and Industry

By adopting the trick of scanning the focal plane through the specimen and catching the changing image upon a vibrating screen we have, it may be claimed, beaten an essential limitation of microscopes—their inability to render visible structure lying in thick sections. Whether the trick of optical scanning in depth will prove useful it is too early to say—all we can claim is that images can be produced in this way which do provide information in depth, and these may be obtained from thick sections of suitable kinds.

The experimental microscopes we have so far built rely on standard lenses, and little further progress is possible without a scaled up lens to give a larger working distance at high aperture. The instrument can be of little practical use until this is available. The images so far obtained are quite good for simple structures, arrangements of small particles such as dust, possibly nuclear tracks in thick blocks of emulsion, and arrangements of fibres. Small particles are seen clearly lying as dots in the image volume, and they may be moved through the luminous block at will by shifting what would normally be the focusing control. There is thus no doubt that the instrument as it stands does give genuine resolution in depth, and that the effect is not merely some kind of illusion.

The contrast is generally rather poor. We are greatly troubled by this, and it is likely always to prove a limiting factor. Contrast may be improved in some cases with oblique or dark ground lighting, but projection then becomes difficult or impossible, which would limit us to direct viewing. In fact, direct viewing, avoiding the screen in place of vibrating elements in the eye pieces, may in any case have advantages.

The quality of image at present obtained is sufficiently good, at a magnification of about three hundred, to show the relative positions of the hairs on the proboscis of insects, and the relative positions of their internal organs, but at this magnification the job would generally be better done with a standard stereoscopic instrument, and we cannot as yet use higher magnifications. Nuclear particle tracks are seen quite well, possibly better than with standard instruments. The principal troubles are the lack of working distance in available objectives which are otherwise suitable, and lack of image contrast even with apparently optimum lighting. The working distance problem waits upon a suitable lens; the impairment in image quality due to lack of contrast might be minimized in the following way, and there are other possibilities.

Consider the apparatus in any of the forms described, but in addition suppose we provide short bright pulses of light at some point in the scan. This would produce a bright layer lying in the block at a position depending upon the position of the pulses along the scan. This bright layer would have high contrast, and it could be moved at will along the *Z*-axis, simply by retarding or advancing the light pulse with respect to the scan. By maintaining some light during the rest of the scan, the structure seen in the normal image plane in the bright layer could be related to structure seen more dimly lying in front and behind this chosen layer. The relative brightness of the rest of the block could readily be controlled, and if the control setting the position of the bright layer on the *Z*-axis were calibrated, measurements in depth could be made quite simply. The brightening can be achieved electrically or mechanically. We are inclined to regard this as an important addition, and it should prove useful not only in improving contrast, by allowing a trading of contrast against detail in depth, but also for making accurate estimates of the position of structures in the *Z*-axis.

The most striking images so far produced are of rough surfaces. Now it happens that the study of rough surfaces is of some importance for industrial control and research into the properties of surfaces and lubricants. It might well prove a useful technique for examining wear, and it could almost certainly be used to advantage for the study of small particles, such as dust and smog, and for nuclear tracks in photographic emulsion.

The images given by arrangements of fibres are quite striking and suggest applications in the fabrics industries, and for such problems as the best bonding materials for brake linings, where it is at present difficult to examine the fibre structure. In medicine, it could be used for tracing the paths of small nerve fibres in muscle (a matter of clinical importance in diagnosing the causes of muscular atrophy) and, if the image can be considerably improved, for studying the connections of the cells of the brain, which is at present very difficult since they are not arranged in thin planes. In general, it could be used for perhaps any application where at present microscopists are tempted to use serial sectioning, and it would have great advantages over serial sectioning which not only is laborious but in some substances impossible, and difficult for any complex structures where the registration between the sections must be maintained.

A further type of possible use is in micro-surgery, a technique used for removing and introducing parts of cells in genetic research, and no doubt applicable for the production of micro-miniature electronic components. It would also be useful for the study of living cells, or groups of cells, for example for studying the effects of radiation.

Man-made electronic components are beginning to approach in scale the components of the nervous system, and many problems and techniques in engineering and biology meet and share common answers. The optical scanning microscope giving a solid image was conceived in terms of biological need, but it might well serve engineers studying micro-structure and producing minute parts. At this stage its uses must be speculative, but we may at least hope that it will in time throw some light in dark places, and make fewer microscopic caverns measureless to man.

I would like first and foremost to thank Professor O. L. Zangwill, for his continuous and generous support and encouragement, even of activities somewhat apart from the main work of his Laboratory. Special thanks are due to my mechanic, Bill Matthews, for his patience, skill and loyalty throughout development work. John Nosworthy provided useful engineering help, and my friends and colleagues Peter Donaldson, Dr H. B. Barlow and Professor D. M. Mackay have contributed vital help, advice and ideas. The Cambridge University Department of Engineering have contributed their skill at high precision machining of special parts. Such progress as we have made would not have been possible without the financial support given by the Human Sciences Committee of the Department of Scientific and Industrial Research.

Bibliography

[1] ALLEN, R. M. *The Microscope*. New York: Van Nostrand (1940)
[2] GREGORY, R. L. and DONALDSON, P. E. K. *Nature, Lond.* (1958), **182**, 1434
[3] GREGORY, R. L. and DONALDSON, P. E. K. *Nature, Lond.* (1959) **183**, 247
[4] GREGORY, R. L. *Proc. Second International Conference on Medical Electronics*, Paris, 1959; *ed.* C. N. SMYTH. London: Iliffe (1960), 591
[5] HOGG, J. *The Microscope*, The Illustrated London Library (1854)
[6] MACKAY, D. M. *Nature, Lond.* (1959) **183**, 246
[7] MARTIN, L. C. *Technical Optics*, Pitman (1948)

RESEARCH ESSAY COMPETITIONS

A list of the other *Waverley Gold Medal* Essay Competition prizewinners and of the Schools Essay Competition prizewinners can be found on p. 452.

THE INTERPRETATION OF COMPLEX SPECTRA

L. BOVEY

Spectroscopy Group, Chemistry Division, Atomic Energy Research Establishment, Harwell

A summary is made both of the atomic designations (multiplicity (S), orbital (L) and total (J) angular momenta) needed to characterize energy levels which give rise to atomic spectra and how these are used to give information on the electronic configurations involved. The methods used to derive such information from experimental investigations of wavelengths, Zeeman effect, hyperfine and isotope structure are described. The relative positions of the lanthanide and actinide series in the periodic table are discussed from a spectroscopic view point. An article by Dr BOVEY on *Techniques for the Examination of Complex Spectra* was published in the September issue of *Research*.

IN THE previous article, *Techniques for the Examination of Complex Spectra*, it has been shown how experimental information concerning spectra of complex atoms can be obtained. With the use of computers, lines in these spectra can be assigned to certain energy level differences within the atom. The interpretation of the spectra involves the finding of these levels, their identification in terms of spectroscopic theory of the atom and their assignment to an electronic configuration. The most important items of information to be derived are the ground state configuration and a series of interrelated levels from which the ionization limit can be accurately deduced.

In a simplified model of the atom, the electrons have associated with them certain quantum numbers. The most important are the principal quantum number, the angular orbital momentum and the spin momentum, denoted by n, l and s respectively. These latter two momenta interact with each other according to definite rules and give rise to the total angular momentum known as j. In a magnetic field, the total angular momentum (j) can take $2j + 1$ orientations with respect to the field. Certain sets of electrons (*e.g.* two s electrons or six p electrons) form a closed group and from a spectroscopic point of view need not be considered further. For a simple atom such as sodium we need only consider the one valency electron producing the spectra. This electron gives rise to a series of levels which are denoted by $^2S_{1/2}$, $^2P_{1/2,3/2}$, $^2D_{3/2,5/2}$ as well as further series. In these levels the superscript is called the multiplicity and represents the effect of the spin of the electron, S, P, D, *etc.* represent the orbital angular momenta (*i.e.* $l=0$, 1, 2 *etc.*) and the subscripts are the total angular momenta (j) which are derived from the joint effect of the multiplicity and orbital angular momenta. If the atom emitting the spectrum is placed in a magnetic field the levels are split further because of the orientation effect mentioned above.

For other elements, when two or more electrons are concerned in the production of the optical spectra, not only is the interaction between the spin and the orbital momenta important but the interaction between the electrons must now be considered. The spins for each electron can be added together as can the separate orbital angular momenta. These two momenta are combined to form the final total momenta for the electrons. Such a form of interaction, technically known as coupling, is fairly common among simple elements and is known as LS coupling. At the other extreme, the spin and orbital angular momenta for each electron can be combined to give the total angular momenta and these individual angular momenta are combined to form the total for all the electrons concerned. This is known as JJ coupling. Capital letters (J, L and S) replace the small ones (j, l and s) for level designation.

With a high number of electrons which may have large orbital angular momenta, the designation for the levels gets more involved although the principle remains the same. Thus, in spectra from rare earths we may expect such designations as $^7F_{0,1,2,3,4,5,6}$ and $^9L_{4,5,6,7,8,9,10,11,12}$.

These levels, which are linked together by the fact that they have the same orbital momenta and multiplicity, form what is known as a multiplet with the series of J values given as the subscript in the multiplet designation. Strictly speaking, such annotation implies LS coupling. In the complex spectra such a coupling is given by only some of the lowest

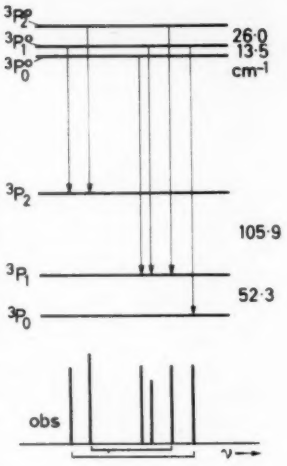

Figure 1. A simple multiplet observed in calcium. The transitions are shown in the top part of the figure and the lines observed underneath. It will be noticed that a line for which the transition is $J = O$ to $J = O$ is not observed. (WHITE[1]. *Courtesy, McGraw Hill*)

levels (*e.g.* the ground state) but they give a conventional designation in which the structure can be discussed. The J value assigned to a level is independent of the coupling and is therefore the main quantum number required from the preliminary analysis. If there is true LS coupling then there are simple relationships between the spacing of the levels and the J values; a simple relationship also holds between the intensity of the lines arising from transitions between such multiplets.

In complex spectra, one cannot rely on these simple relationships and the main rule which can be used is the combination law. According to this, a level with a given value of J, can only combine with another where the change in J (ΔJ) is 0 ± 1; for $J = 0$ the change $\Delta J = 0$ is forbidden. Such a case is shown in the first figure.

Another important consideration is that of parity in the configurations involved. The parity is even or odd according to whether the sum of the orbital momenta of the individual electrons is even or odd. Lines only occur when they involve a transition between configurations of opposite parity. Such an important selection rule is illustrated by lutetium (*Figure 2*). For configurations such as $5d6s^2$ and $5d6s6p$ the parity is even and odd respectively; transitions will only occur between configurations of opposite parity. The configuration $6s^26p$ has odd parity and therefore will not combine with $5d6s6p$ so that no lines can be attributed to such a transition: on the other hand, many lines will be seen resulting from transitions between configurations $5d6s^2$ and $5d6s6p$.

Another general rule is that the strongest lines

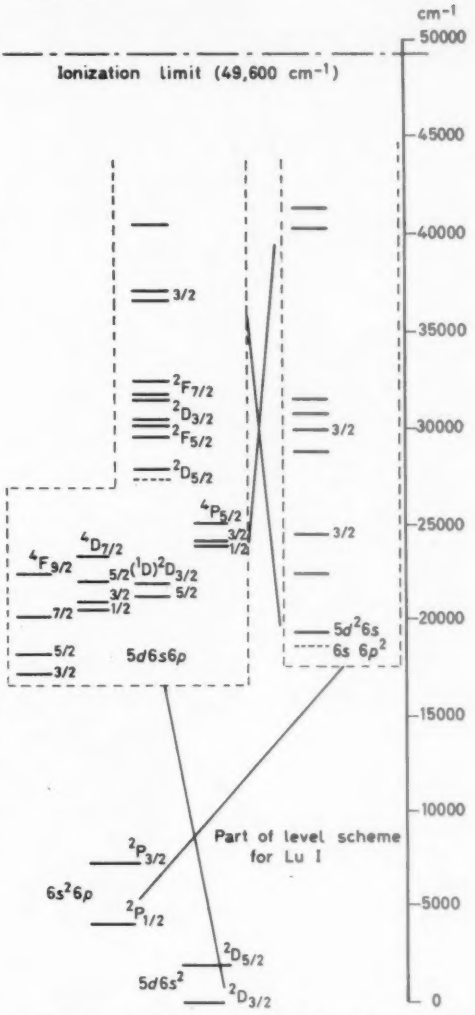

Figure 2. Part of the energy level system of the rare earth element lutetium previously shown more completely as Figure 1 in 'Some Techniques for the Examination of Complex Spectra'. (Research (1960) XIII, 363.) *Transitions occur only between configurations which have opposite parity; thus the configuration 5d6s6p (odd) gives rise to lines combining with the configuration $5d6s^2$ (even[1] and does not combine with the configuration $6s^26p$ which has odd parity.* (KLINKENBERG[2]. *Courtesy Physica, 's Grav.*)

are given when there is no change in the multiplicity of the level so that a ^2P to a ^2D transition would in general be stronger than ^2P to ^4D.

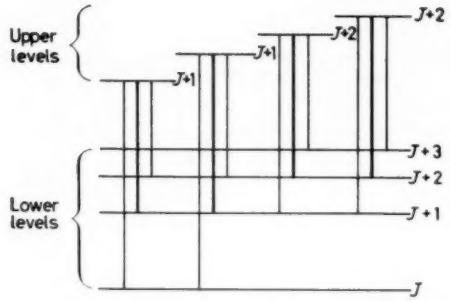

Figure 3. An illustration of the use of lines linking two sets of differences. In the lower set are seen four levels with J values: $J, J+1, J+2, J+3$. The levels J, $J+1$ and $J+2$ will combine with an upper level with $J+1$ value. The heavier middle line links the two differences formed by the lower $J+1$, J levels and the $J+2$, $J+1$ levels. Similarly the heavier line from $J+2$ to an upper level links the two sets of differences $J+1$, $J+2$ and $J+2$, $J+3$

Developments from Experimental Data

As is discussed in the previous article, the first examination of data to find significant differences should produce sets of these which occur more often than would be expected by coincidence. The next step is to see if these differences (each of which is formed by a pair of lines) are linked by having one common line in each of the pairs; if this is the case, the two differences probably occur adjacent to each other and three levels are therefore involved. From these three levels, three lines arise to a fourth level. If these three levels form part of a multiplet, they will have J values differing by one from each other. Levels will therefore begin to be found which have a definite relation between their J values and further examination of levels derived by combinations with different levels will sometimes enable a choice of the J value to be found for reasons of consistency. A simple case of this is shown in *Figure 3*.

In *Table 1* is set out the formation giving rise to a multiplet. At the top are the level values in wave

Table 1. Typical ^7FG multiplet in samarium (ALBERTSON[8]. *Courtesy Phys. Rev.*)

^7F / ^7G	*J* 0 0·00	1 292·58	2 811·92	3 1,489·55	4 2,273·09	5 3,125·46	6 4,020·66
J	22,914·05	22,621·52	22,102·15		*Wave number* (cm^{-1})		
	150 2	80 2	30 2		*Intensities* (*Arc absorption*)		
1 22,914·05	150	150	30		*Intensity* (*emission*)		
	4362·912	4419·332	4523·182		*Wavelength* (Å)		
		23,088·13	22,568·76	21,891·21			
		125 3	60 2	12 1			
2 23,380·72		150	80	25			
		4330·016	4429·664	4566·767			
			23,338·85	22,661·25	21,877·74		
			80 2	30 2	6 2		
3 24,150·80			80	60	20		
			4283·500	4411·585	4569·579		
				32,144·18	22,360·68	21,508·27	
				100 2	100 2	40 1	
4 24,633·74				100	100	40	
				4319·530	4470·886	4648·078	
					32,342·49	22,490·12	21,594·90
					80 1	80 2	5 2
5 25,615·57					80	125	20
					4282·833	4445·153	4629·430
						23,345·89	22,450·72
						100 2	60 2
6 26,471·36						100	50
						4282·208	4452·953
							23,266·92
							300 2
7 27,287·58							300
							4296·743

numbers units of the lower multiplet ^7F and on the left hand side similar data for the upper multiplet ^7G. Where there is a line arising from a transition between two of the levels both the wave number and the wavelength are given together with information on the intensity in various types of sources (arc, absorption and emission furnace). It can be noted that lines for which the change in J is 0 or $+1$ only are observed. When such multiplets have been established, other lines can be forecast and the data examined to see if these are present.

It is usually possible to form a preliminary analysis by such a method using only the wavelengths of the lines, their relative intensity in several kinds of sources and the general principles discussed in the first section. For this reason high wavelength accuracy is the most important experimental criterion for the classification of complex spectra. Other techniques however can be used to confirm and supplement these results.

Zeeman Effect

A very powerful method of determining J is given by studying the spectra produced under the influence of a magnetic field. The light emitted is studied according to its state of polarization. If the light is polarized parallel to the magnetic field the components obtained are called π; if the light is polarized in a perpendicular direction the components are called σ. For two levels combining in which the J values are not equal the number of π and σ components is equal to $2J+1$ and $4J+2$ respectively. If therefore the number of lines can be clearly seen, unambiguous J values for the levels are found directly by inspection. If the two combining levels have the same J value then the number of components is respectively $2J$ and $4J$ (J integral) or $2J+1$ and $4J$ (J half integral). Thus in *Figure 4b* there are nine π and eighteen σ components corresponding to a $J=4$ to $J=5$ transition. In many cases the patterns are not completely resolved into all their component lines but certain general experimental rules can then be applied and data can still be derived from such patterns.

The spacing and intensity of the components also give important information. A constant g (known as the Landé factor) for each level determines the spacing of the components. This factor g is a known function of the quantum numbers involved in the level and for definite forms of coupling (*e.g. LS* or *JJ*) can be calculated. Direct comparison of the spacing of the pattern with that calculated can therefore give confirmation or otherwise that the levels involved belong to certain configurations and types of coupling. The intensity of the lines is determined by the J values and the position of the most intense lines gives supplementary information on J and g.

Series of patterns depending on the relative values of J and g are obtained and for *LS* coupling these patterns have been derived theoretically; some are shown in *Figure 4* for different values of J and g. However, it is also important to realise that the g factors for each level are as characteristic of that level as the J value and therefore if these can be derived from the Zeeman data they give much more information about the level. Moreover in addition to the combination principle which limits values of J, g values for each level must be self-consistent.

Hyperfine Structure and Isotope Shifts

The spin of the nucleus can also influence atomic spectra. For certain isotopes, some of the lines are found to have components. The intensity and number of these lines gives further information about the J values of the combining levels. If a simple case is taken such as that of Pu^{239}, the spin of the nucleus has a value of 1/2. Each level in the

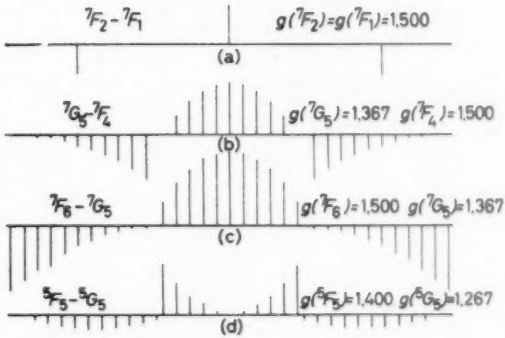

Figure 4. Types of Zeeman patterns. In these four patterns the π components are shown at the top and the σ components along the bottom. The intensity is shown by the height of the line and the relative displacement in the field is given by its position. For different J and g values, different distributions of intensity are shown and these can be used to find the correct multiplicity of patterns derived experimentally. (EL'YASHEVICH[4]. *Courtesy, State Publishing House for Technical Literature (G.I.T.T.L.)*)

atomic spectra can therefore be split into two values, $J+\frac{1}{2}$ and $J-\frac{1}{2}$. In general four lines can arise between two such levels which are split by the influence of the nucleus. If we have J for one level equal to 1 then the total angular momenta are 1/2 and 3/2; if the J value for the other level is 2, total momenta of 3/2 and 5/2 are obtained. Where the nucleus is taken into consideration these total angular momenta are usually denoted by F. These F values are subject to the same type of combination laws as J, *i.e.* the change in $F(\Delta F) = 0 \pm 1$ except when F itself equals 0 when ΔF cannot also equal 0. The ratio of the intensities for two lines arising from one level is given by $(F+1)/F$ if *LS* coupling laws are obeyed. It is found that this coupling is an accurate way of accounting for many hyperfine structure patterns. If therefore the intensity of the lines is determined to a reasonable accuracy, their ratio leads directly to a value of F and thus, if the spin is known, to a value of J. If the total number of possible components is seen then the J values for both the lower and upper states are found. Two such cases for spins of 1/2 and 3/2 are shown in *Figures 5a* and *5b* respectively. In the case where both levels show complete splitting, it is unlikely that all the components would be observed.

When the spectra from two isotopes, say U^{235} and U^{238}, are examined some corresponding lines are found to have slightly different wavelengths. These differences are caused by small shifts in the

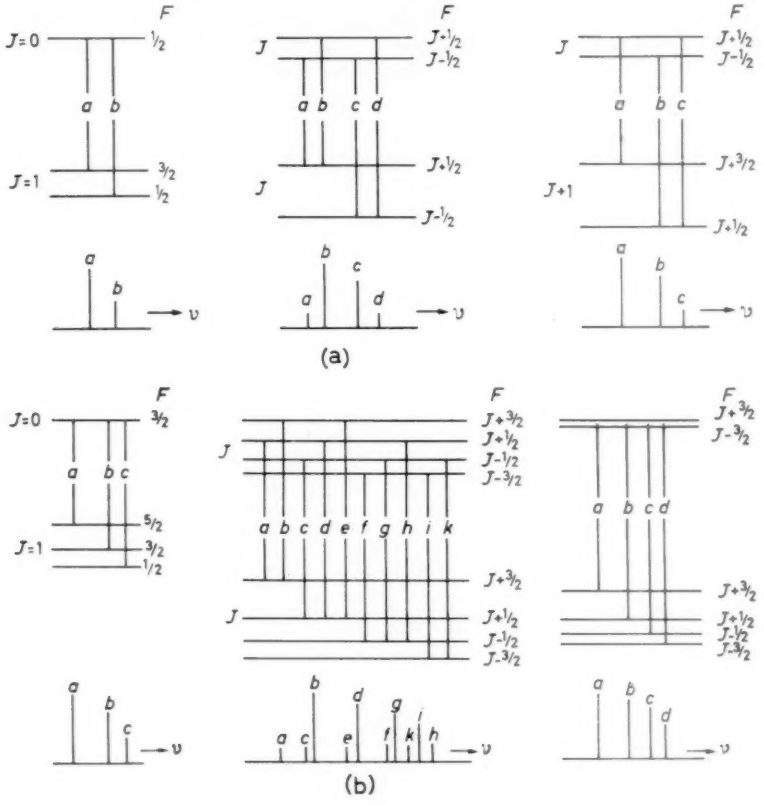

Figure 5 (a). *Simple hyperfine structure patterns when the spin of the nucleus* $(I)=1/2$. *The transitions are shown above and the lines below with the relative intensity shown by the height of the lines.* (b) *Hyperfine structure for the nuclear spin* $(I)=3/2$. *More complicated patterns are obtained than in the previous case but where the splitting in one level is very small the number of components* $=2I+1$, *where I is the spin. This case is shown in the right hand drawing.* (KOPFERMANN[5]. *Courtesy, Academic Press*)

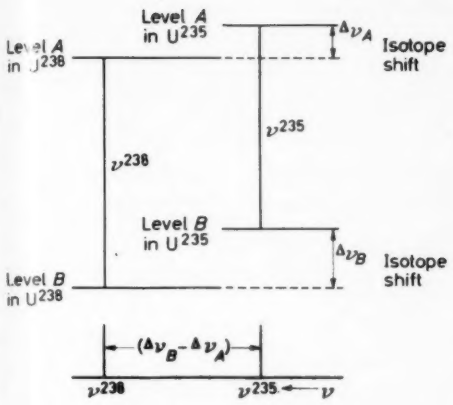

Figure 6. Isotope shifts in uranium. The actual lines observed are displaced because of the isotope shifts in the two levels involved

energy levels which are related to the electronic configurations. The manner in which a shift in the frequency of a line is caused by isotope shifts in the levels is shown in *Figure* 6.

The interaction between the nuclear and electron fields produces both the hyperfine structure and the isotope shift. In general, the presence of hyperfine structure involves a single s electron which passes near the nucleus; the presence of isotope shifts involves a group of penetrating s electrons. Moreover closed groups containing two s and six p electrons do not show hyperfine structure. If therefore lines are found which have only hyperfine structure (which may or may not be completely resolved) the conclusion would be that the levels involved do not arise from the configuration containing closed groups but involve an uncompensated s electron. If both hyperfine structure and an isotope effect were found then the possible conclusion would be the presence of both an uncompensated s electron and a group of penetrating s or p electrons. As in the case of the J and g values, the hyperfine structure and isotope shift effects are constant for any level. Therefore there must be consistency for any levels derived, in that lines which arise from these levels must show the expected isotope and hyperfine structure associated with them. Studies of this structure are experimentally easier for the lanthanide and actinide series as relatively large shifts and pure isotopes can often be obtained.

An example of the use of hyperfine structure and isotope shifts in term analysis is the recent work carried out on americium I and americium II. In the latter spectra of the ionized atom, it was found that many of the lines could be arranged in pairs with a constant difference of 2598·32 cm^{-1}. The levels associated with these terms were found to have $J=4$ and 3. The two levels this difference apart were assigned as $^9S_4$ and $^7S_3$, which had hyperfine structure widths of 1·726 and 1·533 cm^{-1} respectively. The isotope shift of $^9S_4$ was larger than that of $^7S_3$ by 0·042 cm^{-1}. With this established it becomes possible to exclude other coincidental differences near 2598. For example a pair of strong lines with reciprocal wave numbers 27045·50 and 24447·22 (with a difference 2598·28) did not satisfy the condition that the sum of their widths should be 3·259 cm^{-1} and that the difference in the isotope shifts must be near 0·042 cm^{-1}: they did not therefore truly belong to this interval. It was in this way found that the true member of the pair was a line at 24447·05 cm^{-1} (4089·322 Å) which gave the correct hyperfine structure and isotope shift. Some of the pairs of lines making up this 2598 difference are shown in *Figure 7*.

Once the beginning of an analysis has been made by these various techniques, a computer can be used to store the various pieces of information and

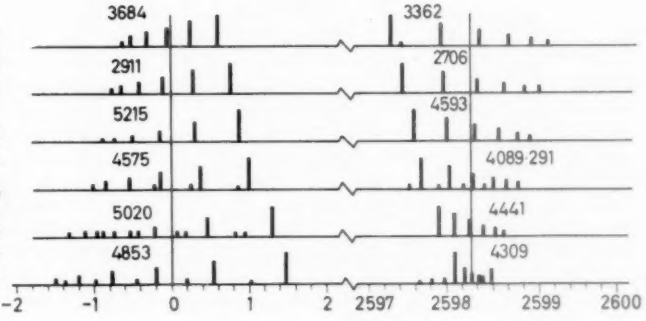

Figure 7. A plot of pairs of lines showing a 2598 cm^{-1} difference in americium II. Various types of hyperfine patterns are present. It can be noted how the lines making up these pairs have hyperfine structure which can be used to verify that the 2598 difference is a genuine one. The line at 4089·291 Å (24447·22 cm^{-1}) is not a genuine member of a pair since its isotope shift and hyperfine structure do not agree with that predicted. A near line at 4089·322 Å (24447·05 cm^{-1}) was found to be the correct member. (FRED and TOMKINS[6]. *Courtesy, J. opt. Soc. Amer.*)

Table 2. Analogies between some ground state configurations of actinide and lanthanide elements

	Actinide		*Lanthanide*	
Neutral Atom	Uranium	U I $5f^3\,6d\,7s^2$	Neodymium	Nd I $4f^4\,6s^2$
	Neptunium	Np I ?	Promethium	Pm I $4f^5\,6s^2$?
	Plutonium	Pu I $5f^6\,7s^2$?*	Samarium	Sm I $4f^6\,6s^2$
	Americium	Am I $5f^7\,7s^2$	Europium	Eu I $4f^7\,6s^2$
Ionized Atom	Americium	Am II $5f^7\,7s$	Europium	Eu II $4f^7\,6s$
	Plutonium	Pu II $5f^6\,7s$	Samarium	Sm II $4f^6\,6s$

* According to some recent work by BOVEY and GERSTENKORN

see if these are consistent with one another. Simple devices such as addition of the values of lines to known levels to find if further genuine levels exist are also possible with internal consistency checks. Theoretically, with high wave number accuracy the whole process could be transferred to a properly programmed machine but general experience in the analysis of complex spectra is also necessary so that a flexible programme, in which intermediate results can be examined, is preferable.

Comparison with Other Elements

The spectrum of any element is related to its neighbours by the fact that an extra electron is added as the atomic number increases by unity. Moreover the two series of spectra with which we have been concerned—the rare earths and the actinide series—have very many analogous properties because in both cases we are concerned with partially filled *f* shells and similar valence electron structure. As can be seen in *Table 2*, americium is in a similar position in the actinide series to that of europium in the rare earths; plutonium corresponds to samarium and uranium corresponds to neodymium.

In both americium and europium the ground states are found to involve seven *f* electrons with one *s* electron in the ionized atom and seven *f* electrons and two *s* electrons in the neutral atom; this similarity of structure is repeated in the multiplets which arise from the electron configurations.

Even however in these cases, resemblances between the higher levels are not very marked and therefore such analogies are only useful in a limited fashion. In such a case as uranium the present analysis of the spectra shows that the similarity between this element and neodymium does not even extend to the ground state configuration.

It is also true that for two neighbouring atoms the number of electrons in the ionized atom of the higher element is equal to that of the neutral atom of the lower element. Thus americium II has the same number of electrons as plutonium I. It is therefore useful to consider whether the term structure of the ionized atom can be carried over to the structure of the preceding neutral atom. Once again these simple considerations are not generally found useful in practice, except in particular instances. Thus in Pu I, the comparison with neighbouring elements shown in *Table* 3 gives several possible ground state configurations *e.g.* $5f^56d^27s$, $5f^56d7s^2$, $5f^67s^2$, $5f^66d7s$ and $5f^77s$ from which the lowest terms can be deduced as ${}^9L_{4\ldots12}$, ${}^7K_{4\ldots10}$, ${}^7F_{0\ldots6}$, ${}^9H_{1\ldots9}$ and ${}^9S_4$ respectively. It will be noted that levels with $J=0$ only occur in the ${}^7F_{0\ldots6}$ term and if lines involving this value for J were found, they should arise only from the $5f^67s^2$ configuration.

Table 3. Possible ground state configurations for Pu I *from consideration of isoelectronic atoms*

Analogous atom	*Configuration*	*Possible ground states*
Two electrons added to U I ($5f^3\,6d\,7s^2$)	$5f^5\,6d\,7s^2$ $5f^5\,6d^2\,7s$ $5f^6\,6d\,7s$	${}^7K$ ${}^9L$ ${}^9H$
Am II ($5f^7\,7s$)	$5f^7\,7s$	${}^9S$
Sm I ($4f^6\,6s^2$)	$5f^6\,7s^2$	${}^7F$

It may well be that when many of these spectra have been analysed certain general rules will be found applicable to all of them and progress has been made to a certain extent in this direction. However, with so few spectra even analysed in a preliminary fashion it has so far been impossible to deduce general rules with any certainty and such help for term analysis is not very great at present. An illustration of experimental similarities in ionized spectra from the rare earth series is shown in *Figure 8*.

It does not seem possible to explain complex atomic spectra purely on theoretical grounds. So far no one has predicted a spectrum which any element would produce in the laboratory. However, some assistance is obtained in the identification of levels by straightforward mathematical theory. It is possible as more spectra become analysed that rules having general application may become apparent and the problem may be simplified.

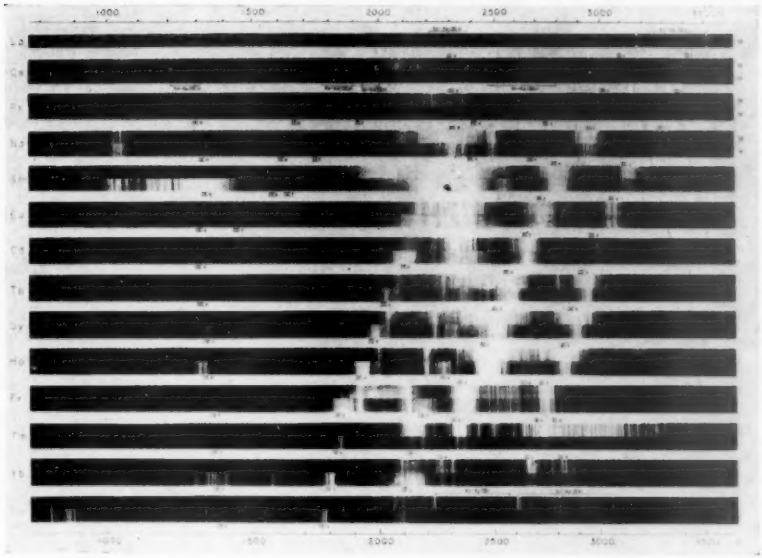

Figure 8. The spectra of the rare earth elements taken in the 1000–3500Å region. The regular shift in wavelength of certain configurations can be observed. (DIEKE, CROSSWHITE *and* DUNN.[7] *Courtesy, J. Opt. Soc. Amer.*)

Some help in classification may be given by comparison with spectra obtained from crystals. Because of the special conditions in the crystal, these spectra are somewhat simplified and in any case involve the more highly ionized spectra (in general the neutral atom with three electrons removed). In practice it has been found that the reverse situation is usually required; that is in order to explain the crystal spectrum, that of the free ion should be known. Theoretical work on the calculation of levels in crystal spectra is in a more advanced state than corresponding work on the free ion mainly because of the simplification introduced when the ion is in a crystal field.

Conclusion

Both the techniques for observation and methods for interpretation of complex spectra can be developed to overcome the difficulties involved. Most rapid progress can be made when they are used simultaneously. It is also true that advances in the study of isolated spectra, whilst interesting in themselves, do not really allow a general advance. It would therefore seem that it will be a considerable time until there is agreement with the statement made in the preface of the book (published in 1935) *The Theory of Atomic Spectra* by CONDON and SHORTLEY, that 'the interpretation seems to be in a fairly closed and highly satisfactory state' if this phrase is taken to include the spectra of the actinide and lanthanide elements.

References

[1] WHITE, H. E. *Introduction to Atomic Spectra* New York: McGraw Hill, 1934

[2] KLINKENBERG, P. F. A. *Physica, 's Grav.* (1955) **21**, 53

[3] ALBERTSON, W. *Phys. Rev.* (1935) **47**, 320

[4] EL'YASHEVICH, M. A. *Rare Earth Spectra* Moscow: State Publishing House for Technical Literature (G.I.T.T.L.), 1953

[5] KOPFERMANN, H. *Nuclear Moments* New York: Academic Press, 1958

[6] FRED, M. and TOMKINS, F. *J. opt. Soc. Amer.* (1957) **47**, 1076

[7] DIEKE, G. H., CROSSWHITE, H. M. and DUNN, B. *J. Opt. Soc. Amer.* In preparation

MAN POWERED FLIGHT

T. NONWEILER

Department of Aeronautical Engineering, The Queen's University, Belfast

The offer of a £5000 prize for a sustained and controlled flight powered by the muscular activity of the aircraft's occupants, has attracted attention recently to the efforts now being made to fulfil man's primeval urge. Some accomplishments of past aircraft, and the limitations of those likely to be made, are here discussed, with special mention of man's ability to produce power. The helicopter and the ornithopter have their protagonists, as is pointed out, whilst others pin their faith in an adaptation of the fixed-wing sailplane, with man supplying the motive power to a propeller.

WE HAVE been brought up in the unquestioning belief that man cannot fly; perhaps for no better reason than the feeling that, if God had intended man to fly, He would have given him wings. Certainly man's attempts to fashion his own wings and consign himself to the winds have not often met with encouraging success. The history of flight is filled through all recorded time with stories of leaps to an unintended suicide, and if legends are any guide, such exploits took place in pre-historic ages as well.

In more recent times, there has admittedly been a different approach. Lilienthal's classic experiments with hang-gliders just before the turn of the century, which contributed so much to the understanding of wing design, depended to some extent at least on muscular power, in as much as he made a running take-off; and, of course, nowadays the art of gliding gives some satisfaction to those who relish the sense of personal conquest that comes from flight without engines. If man's own exertions could be used to replace the powered launch, not only would the sense of personal accomplishment be considerably enhanced, but the tiresome formalities associated with a fixed aerodrome and take-off routine would be avoided. This might be a reasonable goal to seek in our efforts to produce a man-propelled aircraft, but we are still some long way from its attainment.

There have been many direct attempts to replace engines by muscle, ever since powered flight became a reality. Just before the first world war, and again after it, there was great activity in France. Spurred on by the offer of prize money, a number of 'aviettes' were constructed and flown—and the money duly won for various modest accomplishments. These machines were essentially bicycles to which were geared, or otherwise attached, flapping wings. The 'flights' were precipitated by a headlong ride towards an inclined ramp, and the consequent leap was protracted to a greater or less extent by the flapping of the wings: the whole performance could be considered a warm-weather ski-jump, in which sport, incidentally, the aerodynamic lift-force developed by the skier plays an important part. Another prize offered in Germany in the 'thirties once again stimulated development of muscle-powered machines, this time with more interest evoked in fixed-wing aircraft than in ornithopters. The best known result was a propeller-driven aeroplane designed and built by Haessler and Villinger at Frankfurt (*Figure 1*), the only detraction from its complete success lying in the manner of its take-off: it was launched from a rail by the impetus of a stretched-rubber catapult. Flights of up to 40 seconds duration, and 400 yards in length were recorded, and the aircraft reached a height of 30 feet—figures which all suggest that the performance depended to a larger extent on the energy of the pilot, transmitted by foot pedals to the propeller, than on the impetus of the catapult.

It appears, however, that the most successful of the pre-war machines was that designed by Bossi and Bonomi in Italy. This aeroplane—called the 'Pedialante'—outwardly resembled a high-performance sailplane except for the presence of its two tractor propellers, mounted on the wings, which were again driven by the efforts of the occupant transmitted through foot pedals (*Figure 2*). It was reported at the time that the 'Pedialante' flew 1000 yards on a circular course, but it has only now been pointed out by Signor Bossi—recently 'discovered' in the United States—that this was after a *self-propelled* take-off. The machine was driven along the ground in the manner of a bicycle and the propellers were only brought into use when flying speed was reached. Thus this aircraft seems to have been the first which can truly be said to have achieved man-powered flight, without any shadow of doubt cast by the use of energy stored up prior to flight.

Recent Developments

Since the war, interest has been re-awakened, particularly in Britain, where two aircraft have been built, although at the time of writing, neither has

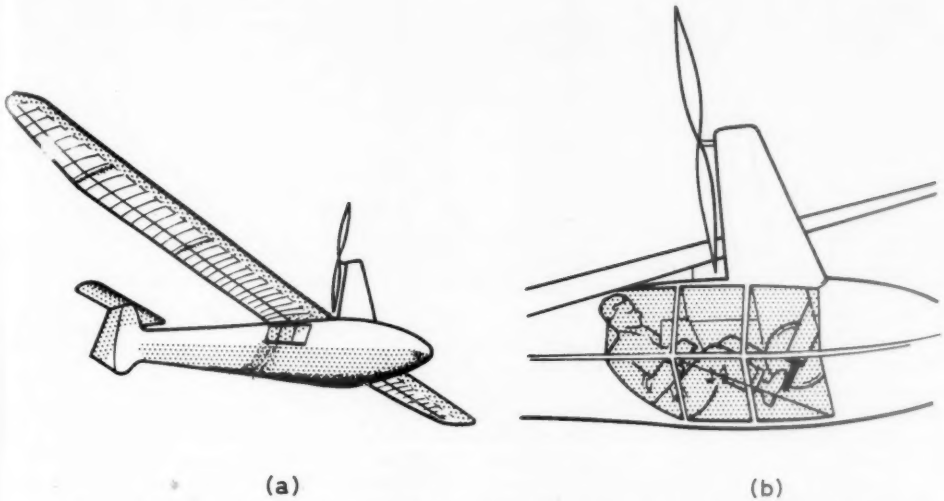

Figure 1. (a) *The Hessler-Villinge machine;* (b) *shows the crew member position and the twisted belt drive* (*Courtesy Aeronautics* (*1959*) **41,** 24)

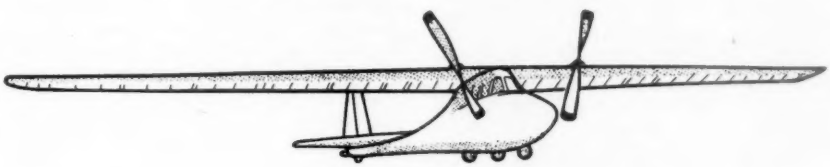

Figure 2. The Bossi-Bonomi 'Pedialante' (*Courtesy Aeronautics* (*1959*) **41,** 24)

been flown. One of these is an ornithopter, designed by a South African sculptor, Emil Hartmann*; and the other is a fixed-wing machine constructed at Cardington by Perkins, who has pioneered the development of inflatable rubber wings (*Figure 3*). He has thoroughly mastered this novel technique of construction and can now produce the most beautifully streamlined wings, whereas those of us with less experience might have to be content with a bulbous and ill-proportioned shape the result of building a structure round a number of blown-up sausage balloons! There are a number of other aircraft in the design stage—including a twin-seat wooden aeroplane with which the author has been associated for the last five years (*Figure 4*)—and again the endeavour has been given impetus by a competition for a cash prize of £5000. This has been offered by Henry Kremer, and its terms have been drawn up by the Royal Aeronautical Society*; it will, no doubt, attract a lot of competitors, not only (it is to be hoped) for reasons of pure avarice. The prize will be awarded for the first flight in a figure of eight about posts half-a-mile apart, without the use of stored energy (as for instance in a rubber catapult), and many people feel this will be an extremely difficult task. There is, to boot, a ten-foot hurdle to cross at the starting line, which (with cruel cunning!) also becomes the finishing line as well.

In other countries, too, there are aircraft being built. In Russia the emphasis seems to be on ornithopters, with an approach very wisely through the development, first of all, of machines powered by small engines. Work of a similar kind is being financed by the Rockfeller foundation in the United States. It is pleasant to see this surge of interest in a modest and primitive urge, in days when other products of aeronautical technology invite attention only when travelling higher, faster or further than ever before.

* See *Research* (1960) **XIII**, 329

Biological Limitations

Although historically, therefore, there is evidence that man *can* fly, and that a number of people are now sufficiently convinced of this that they are spending their time in re-proving it, nonetheless man-powered flight still appears only a marginal possibility. There is no suggestion yet that we shall soon be able to gambol with the gulls if the fancy takes us, and for the less energetic of us the margin of possibility will be too small to be a temptation. Biologically we are not well designed for flight, and the machines we construct can do little to remedy the deficiency which is a direct product of a poor power–weight ratio.

This appears to be yet another ramification of the square–cube law: broadly speaking, an animal's power is not so much limited by the mass of its muscle but by its rate of intake and absorption of oxygen—in other words by the rate at which it can burn up fuel. It would be too coarse a simplification to suggest that this is proportional to the surface area of the lungs, but it does happen to work out that the continuous power developed by various forms of animal life varies roughly in proportion to their weight raised to the two-thirds power—or since the specific gravity of all animals is about the same, in proportion to the two-thirds power of volume, which is dimensionally an area. The power needed for flight, on the other hand, tends to increase more rapidly than the weight, since heavier animals must fly faster—supposing at least that they are geometrically similar. One sees tangible support for this reasoning in the fact that the largest birds are flightless. Thus if it is difficult for man to fly, he can take comfort from the fact that it must have been still more difficult for Pegasus, and even more so for Disney's winged jumbos.

Power Production

The only firm deduction that can be made from such reasoning is that it is unwise for man to try to fly by shaping himself as a giant replica of a bird: geometrical similarity is not only unnecessary but undesirable—appreciation of this fact might have saved many erstwhile bird-men their lives. Also, it needs to be pointed out that not *all* of man's available energy depends on oxygen supply. The usual method of energy production certainly stems, in the final analysis, from oxidizing lactic acid to preserve a supply of glycogen within the muscles. But certain reserves can be called upon, derived from the hydrolysis of chemical stores, during which process lactic acid accumulates in the muscle concerned and ultimately produces the pains of fatigue. These reserves, incurring the so-called 'oxygen debt'—which must be repaid by heavy

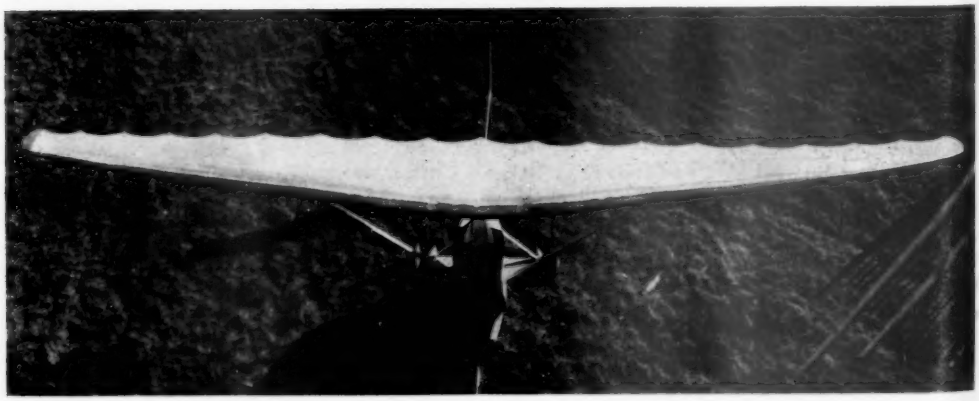

Figure 3. Perkins inflatable wing machine. (*Courtesy 'Aeroplane'*)

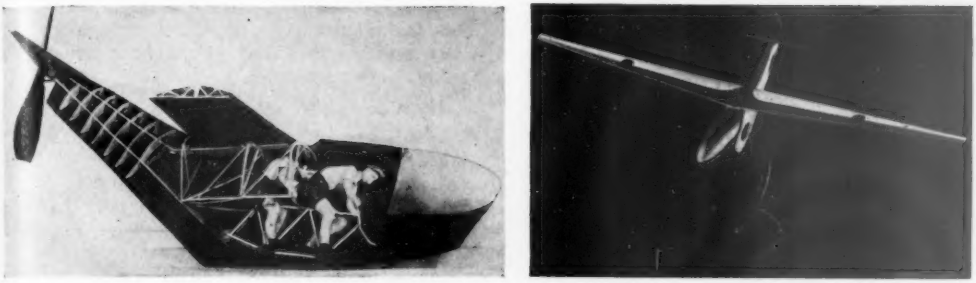

Figure 4. Models of the author's twin-seat pedal machine

respiration during recovery to oxidize the accumulated acid—depend on the amount of muscle which can be brought to work. Thus in certain activities, like cycling, these reserves allow a man to produce a power of 1½ h.p. for perhaps 10 seconds, before fatigue sets in; whereas if he must cycle for a long time without becoming fatigued, then he dare not incur the oxygen debt and would be limited at most to an output of about ½ h.p. On the other hand, if, for example, only the arm muscles are used, although he might still produce ½ h.p. over a long period, he could not greatly improve on this figure even for short durations, since he has relatively little in reserve.

Use of these reserves of energy allows man to overcome the handicaps of his size and achieve flight for a short period before fatigue overcomes him. By tying their wings to their arms, the early bird-men denied themselves this advantage. By gearing their machines to a bicycle, the latter-day aeronauts have achieved brief glory. Many means of power-production other than cycling have been suggested which holds the hope of even greater trade upon the oxygen debt: an investigation by Oscar Ursinus during the pre-war craze for muscle-flight in Germany, showed that of all the various possibilities, combined foot and hand cranking was the most energetic. But if the power required for flight could be brought down to that level which man can develop *continuously*, the method of power production would be largely immaterial. However, the other alternative—to raise the level of this maximum continuous power—appears more readily accomplished, as it is known that this is the effect of inhaling oxygen-rich air. Equipped with an oxygen set, it may well transpire that man can fly for as long as the capacity of the gas cylinders permits.

Design

Except for this device, there seems little one can do to improve the human engine, although obviously there will be a great variability in performance from one man to the next. For the aeronautical engineer, this is part of the fascination of the whole endeavour, for in the history of powered flight one has to admit that progress, as measured numerically in speed, height or range, has largely been the result of bigger or better aero-engines: it has not always been so easy to disentangle the effects of improvement in aerodynamic and structural design.

It is difficult to separate these two aspects of design, since improvement in the one is usually only obtained to the detriment of the other. Thinking in terms of fixed-wing aircraft, one would logically design the aircraft to be capable of being flown for the least possible power, and this usually results in a shape of aircraft resembling a glider, which has rather similar requirements. Wide wing-span, and careful streamlining, pay handsome dividends in performance, despite the expense in structural weight. Unlike the glider, one can make a virtue of the necessity of the short duration of flight of the man-powered aircraft by envisaging that it should always fly close to the ground, where it rides more easily on a 'cushion' of air; and the proportions of the aircraft can be arranged to suit. Likewise, one recognizes that flight in gusty conditions or at very high speeds will be impossible, and one can employ special highlift wing-sections, to provide benefits which would otherwise have to be eschewed in favour of more conservative design.

But these are relatively elementary considerations: any important aerodynamic innovations will perhaps be those directed towards the avoidance of turbulent flow over the aircraft. The suction of the

air through slots on the wing is a technique already well-known in this connection: attention, too, has lately been drawn to the elastic properties of the skin in damping out turbulence. However, it would be unwise to expect any surprises, as anything which improves the performance of the man-powered aeroplane could also be applied to gliders, and *vice versa*, and any good ideas can be expected to be exploited more readily in that highly competitive sport, backed as it is by commercial investment.

The fuselage design admittedly presents a different problem from that of the glider, since one must allow for the correct posture of the hard-working pilot. There is too the problem of a transmission system: with such low powers light-alloy chains are not only very light but highly efficient, and tolerant of a surprising amount of twist.

In the structural design, likewise, one can take advantage of the limited flight duration by not strengthening the framework against the hazards of all-weather flying. Gliders are built, for instance, so that their wing spars can carry a weight of nine times that of the aircraft before they fail, so as to withstand the large accelerations met (say) in seeking thermals inside thunder clouds. The man-powered aeroplane is not likely as yet to achieve enough height to get itself into bumpy weather, and such precautions are an unnecessary luxury. Thus one can allow the spars to fail when carrying perhaps three times the weight, and relieve the occupants of a considerable burden of weight. Carrying out pruning like this right through the structure allows the weight of the aircraft to be reduced to between a half or a third of that of a glider of comparable size—at the expense of air-worthiness, but, of course, with profit to the performance. Thus, although man could not develop enough power to fly a conventional glider (about 3 h.p. is needed for a single-seater, in the absence, of course, of aid from air currents) he can fly an outwardly similar, but internally pruned, version of the same aircraft. There are two limits to the amount of pruning which can be done: firstly, one finds that the structure becomes so elastic that stiffness rather than strength becomes a criterion; secondly, with a wooden or metal construction the thin sheets of material one wants just cannot be made. These are indeed unusual problems, and innovation in the materials of construction—perhaps with plastics, or inflatable rubber construction —could bring about worthwhile improvements. But, of course, if these, and any aerodynamic improvements, enable the man-powered aeroplane to be flown high enough to find itself in rough air, one would be forced to put the strengthening one has taken out back. The performance would then, of course, suffer, and one sees in this situation the vicious circle separating the short man-powered hops now possible, and the ultimate dream of a link-up with sailplaning.

Other Possibilities

Whilst this division exists, one is tempted to turn to other configurations of aircraft in the hope of extending the short hops. The helicopter springs naturally to mind, as it might be thought that with its ability to rise from the ground so slowly, it would need less power and use up less of the precious human reserves. But this is not so. Even with 100 ft diameter blades, a single man helicopter uses about 2 h.p. in hovering close to the ground—power which is, of course, thrown away into the downwash of the slipstream. At forward speeds of between 10 and 20 m.p.h., the power required for flight is a great deal less, but it still does not appear to be any smaller than that needed to propel a fixed-wing aircraft at perhaps double the speed.

There are many protagonists of the ornithopter, but all too often they base their support on specious reasoning. Whatever the idealist may feel about the lessons the birds can teach us—and after all their prowess comes chiefly from the extreme flexibility of their wings, rather than the manner of their wing beats—and however satisfying it may be to combine the tasks of lifting and thrusting in one apparatus, there is no doubt we know little about the engineering or the aerodynamics of flapping flight, and we are not in a position to judge its merits with any certainty. All that can be said is that at the speeds of fixed wing aircraft, the propeller is a very efficient device—one designed for a limited range of flight speed might waste less than 10 per cent of the power put in to it. It would be very difficult to improve on this, particularly bearing in mind its mechanical simplicity. Again, in hovering or very low speed flight, a lot of energy is wasted in obtaining the reaction on the air, and the ornithopter, like the helicopter, although quite capable of providing lift and thrust in such a condition, only does so at the expense of considerable power. Indeed, like the helicopter, it is a device which holds most promise in relation to flight at speeds of perhaps a half or more of those characteristic of fixed wing aeroplanes. Whether at these speeds it offers the possibility of flight at a smaller expense in power, and even if it does, whether this power advantage is sufficient compensation for the added structural weight of a divided and hinged wing, are questions which it is impossible to answer without experiment. And experiment is certainly worthwhile.

PROPERTIES OF GRAPHITE

Part II - Thermal and Electronic Properties

L. C. F. BLACKMAN

Department of Chemical Engineering, Imperial College of Science and Technology

The thermal and electronic properties of near-ideal and defective graphites are discussed in terms of the fundamental anisotropic crystal structure. The first part of this tripartite article dealing with the preparation and the structural and mechanical properties of graphite was published in the October issue of *Research* and Part III on crystal compounds and irradiated graphite will appear next month.

THE distance between carbon atoms in the plane of the hexagonal layers of graphite is 1·415 Å compared with the interlayer spacing of about 3·4 Å[1]. As a result of this pronounced difference in the bonding between the atoms it is possible, to a first approximation, to discuss the properties of graphite in terms of the single hexagonal layer. However, for more refined treatments it is necessary to use a three-dimensional model to account for finite interlayer interactions. In some cases, for example, the variation at low temperatures of the electrical resistivity, the two dimensional approach is unsatisfactory, but in other cases the general predictions of the two models are similar. As stressed in *Part I*[1], structural and chemical imperfections lead to pronounced changes in the properties of graphites relative to the ideal solid, and they can influence the choice of model in particular cases.

Thermal Properties of Graphite

It is normal to think of the thermal properties of graphite in terms of its unique position as a high temperature material. Recent developments in rocket missiles, ram jets and high power nuclear reactor heat sources require materials capable of operating at temperatures of 2000°C and above, and the future of graphite for such purposes seems promising. A novel high temperature use of synthetic pyrolytic graphite, according to the commercial literature of the *Raytheon Company*, U.S.A., is as the leading edge of space vehicles for re-entry into the atmosphere, where a high thermal stability coupled with good thermal conductivity is desired. However, not all applications of graphite utilize its high temperature properties. At normal temperatures its relatively high and controllable thermal conductivity, its ease of fabrication and its low density makes graphite an attractive material for many purposes quite apart from its extensive use in nuclear reactors.

In this article the main thermal properties of near-ideal and defective graphites are described in terms of the fundamental properties of the ideal crystal lattice.

Specific Heat

The room temperature specific heat of graphite is well below the value of 6 cal/g atom-deg (or $3R$ where R is the gas constant) predicted by the classical Dulong–Petit law. Following the early low temperature measurements of NERNST[2], MAGNUS[3] in 1923 measured the specific heat of graphite in the temperature range 44·1° to 1100°K. Even up to the highest temperature the specific heat continued to rise but remained below the Dulong–Petit value. Later measurements by HOVE[4] revealed a plateau at around 6 cal/g atom-deg extending from about 1750° to 3500°K. Above this the specific heat increased exponentially (with an activation energy of ~7 eV) to greater than 12 cal/g atom-deg. This increase, which is attributed to the onset of lattice vacancies or other forms of thermal defects, is analagous to the 'pre-melting' increase observed with many solids. Measurements down to 1°K have been made by DESORBO and NICHOLS[5]. The collected data from various investigations[2–6] over the complete temperature range to ~4000°K is plotted in *Figure 1*.

Variation with Temperature

An interesting feature of the specific heat of graphite is its variation with temperature. Although even cubic metals show deviations from the classical prediction that at low temperatures the specific heat of solids should vary as T^3, graphite exhibits a marked deviation by varying as T^2 until very low temperatures. For a long time it was thought that T^2 variation was a direct result of the two-dimensional nature of the graphite lattice, but it appears, however, that the mere existence of an anisotropic crystal structure does not automatically cause the anomaly.

Below ~10°K near-ideal graphite exhibits a T^3 variation of specific heat in agreement with theoretical models[5]. For such graphite the heat capacity can be expressed

$$C_v = aT^3 + \gamma T$$

where the small linear term is the electronic specific heat. The constant γ is proportional to the density

of electron states at the Fermi surface and is estimated[5] to be $\sim 0{\cdot}6 \times 10^{-5}$ cal/g atom-deg. The main contribution to the specific heat therefore arises from the thermal energy of the lattice vibrations. The T^3 dependence is to be expected when the interlayer vibrations dominate relative to out-of-plane vibrations. Above 10°K the principal contributions are from these out-of-plane modes of vibration (see ref. 7 for further references).

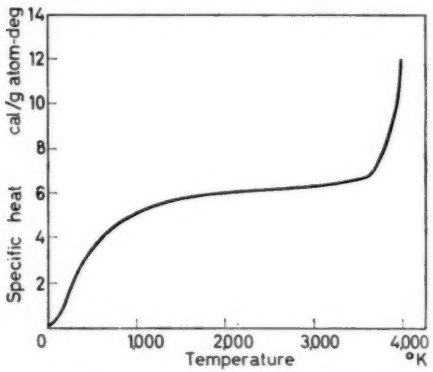

Figure 1. Representative plot of the specific heat of graphite[2–4]

For imperfect graphites the T^2 variation persists to the lowest temperatures of observation. This behaviour is attributed to a marked decrease in the out-of-plane elastic constant c_{44} (increased interplanar shear resistance) due to the lack of three-dimensional order[1]. As a result of this the bond bending vibrations of the layer planes can be excited at much lower frequencies. BOWMAN and KRUMHANSL[7] postulate that this effect might also be the origin of the observed higher specific heat of imperfect graphites relative to the near-perfect crystal. Certainly the difference in the specific heat appears to be structural in origin since DESORBO and NICHOLS[5] have shown that changing the electronic structure of natural graphite, *via* substitution of 0·04 per cent of the carbon atoms by the electron acceptor atom boron, has no perceptible effect on the specific heat in the liquid helium range.

Debye Characteristic Temperature

By comparing the data of *Figure 1* with the equations of the Debye theory of specific heat it is apparent that the Debye characteristic temperature Θ (which is a function of the velocity of sound in the solid and of the interatomic distance) of graphite is very high. NERNST[2] in 1911 suggested that it was necessary to ascribe two Debye temperatures to graphite to account for the in-plane and inter-plane modes of vibration. MAGNUS[3] estimated $\Theta_a = 2280°$ and $\Theta_c = 760°$K. More recently, KRUMHANSL and BROOKS[8] interpreted their experimental data in the range 15° to 1000°K in terms of $\Theta_a = 2500°$ and $\Theta_c = 900°$K.

Thermal Conductivity

It is generally not realized that the room temperature thermal conductivity of highly crystalline natural and synthetic graphite in the direction of the layer planes exceeds that of the best metal conductors; it is, in fact, second only to diamond on an absolute scale. In metals, which are generally good conductors of both electricity and heat, the principal carriers of the heat are the electrons in the conduction band (roughly 100 times more effective than the lattice vibrations). In graphite, even though it possesses an appreciable electronic conductivity, it is the lattice vibrations that determine the thermal conductivity. If the Wiedermann–Franz law is assumed to hold then the electronic contribution to the thermal conductivity of artificial block graphite is estimated[9,10] to be only about 0·5 per cent at room temperature and about 6 per cent at 1000°K. Values for natural graphite are reported to be somewhat less[9]. This general result that phonon–electron collisions have a negligible effect on the thermal conductivity has been directly confirmed by BOWMAN, KRUMHANSEL and MEERS[11] who observed that substitution of 0·1 per cent of the carbon atoms in pyrolytic graphite by boron (which would have little effect on the lattice vibration frequencies) had no effect on the thermal conductivity, although the electronic conductivity was increased by 20 per cent (see later).

It is experimentally extremely difficult to obtain accurate measurements of the thermal conductivity of very small single crystals. Nevertheless BERMAN[9] reports the room temperature thermal conductivity of a flake of natural graphite in the directions parallel and perpendicular to the preferred *a*-axis direction as 2·5 and 0·8 W/cm-deg respectively. The lower '*c*-axis' value is in accordance with the lower Debye temperature for this crystal direction. Other values obtained from samples of natural[12] and synthetic pyrolytic[11] graphite having presumably a higher degree of preferred *a*-axis orientation exceed the value for copper (3·88 W/cm-deg) (*cf. Figure 2*).

Polycrystalline graphites exhibit lower thermal conductivities than single crystals due to a combination of effects involving the finite particle size, lattice imperfection and crystal boundary effects, and the presence of macroscopic pores. However, highly graphitized material is still a good thermal

conductor (*cf. Figure 2*). As with other properties, artificial graphites prepared by extrusion or pressing show an anisotropy of thermal conductivity related to the degree of preferred orientation of the crystallites[13]. This is illustrated in *Figure 2* which plots the temperature variation of thermal conductivity for a number of graphites of different origin and preferred orientation. The position of the maximum is related to the average crystallite size[9,14].

The low temperature thermal conductivity of artificial and natural graphite of small crystallite size varies with temperature as $\sim T^{2\cdot3-2\cdot8}$ whereas the specific heat of such material varies as T^3. The dependence on a power higher than two is generally attributed to the predominant influence of non-graphitic intercrystalline matter. HOVE and SMITH[14] show that if this matter has a poorly crystalline diamond-like structure and behaves as a discrete isotropic Debye solid (with a dependence on T^3) then the observed temperature variation of the two-phase material can be explained. Striking evidence in favour of this hypothesis is the T^2 variation of the low temperature thermal conductivity of highly crystalline graphites (*Figure 2*).

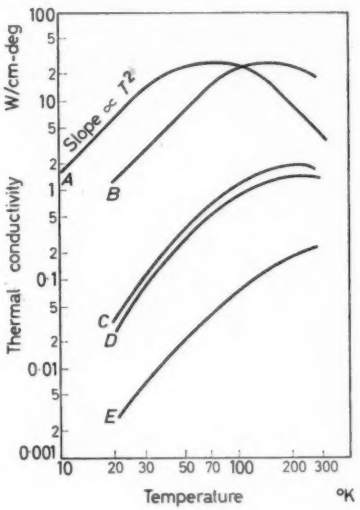

Figure 2. Thermal conductivity of different graphites:

A—Canadian natural graphite[12];
B—synthetic pyrolytic filament[11];
C—artificial block graphite, parallel to extrusion axis[13];
D—artificial block graphite perpendicular to axis[13];
E—graphitized lampblack[13]

The variation of the thermal conductivity of artificial polycrystalline graphite in the high temperature range to $\sim$ 3800°K has been determined by EULER[15] and HOVE[4] using quite different methods. A pronounced exponential decrease, with an activation energy of $\sim$7 eV, occurs at around the same temperature as the increase in specific heat mentioned above. Doubtless both phenomena are related to the formation of thermal defects.

A useful empirical observation (*cf.* CURRIE, HAMISTER and MACPHERSON[16]) is that for a wide range of graphites the product of the room temperature thermal conductivity in cal/sec-cm-deg and the electrical resistivity in ohm cm is about 0·00031. This allows a fairly accurate assessment of the thermal conductivity of a graphite from the value of the more easily measured electrical resistivity. From what has been said above, this relationship between electrical and thermal conductivity (which is not related to the Wiedermann–Franz law as this applies only to purely electronic conduction) obviously does not hold for chemically doped graphites where the electron energy band structure is altered.

Thermal Expansion

The thermal expansion of well ordered graphite is extremely anisotropic as shown by the careful x-ray diffraction measurements of NELSON and RILEY[17]. In the *c*-axis direction the mean expansion coefficient over the temperature range 15° to 800°C was found to be $28\cdot3 \times 10^{-6}$. In the *a*-axis direction the expansion is complex. Up to 150°C the lattice contracts slightly ($\sim -1\cdot5 \times 10^{-6}$); this is followed by a region of zero expansion until finally a small positive expansion ($\sim +0\cdot9 \times 10^{-6}$) occurs above about 600°C. The initial contraction, which is also observed in the hexagonal crystal of zinc, has been attributed to a Poisson ratio effect, but this is not entirely a satisfactory explanation and the matter cannot be considered as settled at present.

Other workers using graphites of different degrees of crystal perfection have established that the interlayer expansion is relatively insensitive to the degree of crystal disorder. It is surprising, therefore, that bulk polycrystalline graphite exhibits only a small thermal expansion relative to that of the constituent crystallites. Typical values for reactor block graphite are $1\text{–}2 \times 10^{-6}$ in the direction of the preferred orientation of the crystallites and $3\text{–}4 \times 10^{-6}$ in the perpendicular direction. The

reason for the low expansion is not well understood at present, although it is presumably associated with the presence of pores in the material.

A technically useful feature of bulk polycrystalline graphite is that the dimensional change with temperature is usually quite smooth up to about 3500°C[4]; also, graphites of different origin tend to behave similarly up to the highest temperatures although their absolute expansion coefficients may differ. This means that the expansion at high temperatures may be assessed, to a first approximation, by extrapolation from room temperature measurement[16].

Thermodynamic Properties

Quite large variations exist in the literature concerning the high temperature thermodynamic properties of graphite. This is to be expected in view of the extreme technological difficulties involved in making accurate measurements at temperatures as high as 3000°C or above.

At atmospheric pressure graphite sublimes directly into the gaseous state at around 3500°–3700°C. The vapour pressures at lower temperatures are listed in *Table 1* (after KELLY[18]).

Table 1

Vapour pressure mm Hg	*Temp.* °C	*Vapour pressure* mm Hg	*Temp.* °C
10^{-5}	2129	10^{-2}	2681
10^{-4}	2288	10^{-1}	2926
10^{-3}	2471		

The triple point has been reported by BASSET[19] to occur at 105 atmospheres and 4000°K. The low vapour pressure of graphite up to about 2200°C makes it an excellent high temperature material for use in non-oxidizing atmospheres.

The heat of sublimation of graphite has not been measured by a direct calorimetric method, but a value of 170·4 kcal/mole seems likely from various lines of experimental evidence (*cf.* KERN[20]).

Graphite–Diamond Transformation

The transformation of graphite into diamond has been attemped by many investigators during the past 140 years, but the first authentic synthesis is believed to be that of H. T. HALL[21] on the 16th December 1954 at the General Electric Research Laboratory, New York.

Basically the problem, apart from exceptionally difficult technology, is one of reaction rate *versus* thermodynamics. At ordinary pressures graphite is the stable form of pure carbon (for example, when diamond is heated to 2000°C it is completely converted to graphite) whereas diamond is the stable form at pressures around 30,000–100,000 atm. at temperatures in the range 1000°–3000°C[22]. However, in practice diamond failed to be produced from pure graphite under the extreme simultaneous conditions of 200,000 atmospheres and nearly 5000°C[21]. The reason for this was attributed to the very slow reaction rate imposed by the pressure, and accordingly attempts were made to catalyse the transformation. This proved successful; possible catalysts are tantalum, nickel or some other special metal or alloy, which under the conditions of operation melt, expand and percolate into the mass of graphite. Full details have not yet been published concerning the starting materials, the experimental procedure and the new thermodynamic data that has been obtained.

Electronic Structure of Graphite

The bonds in graphite are of two types; σ bonds which are localized between adjacent carbon atoms within a layer and π bonds which are represented by orbitals extending over the complete layer. It is the mobile π electrons which give graphite its characteristic metallic-type properties, and for this reason most discussions of the electronic structure of graphite ignore the contribution of the σ electrons. The discrete energy levels of the individual π electrons are replaced in the infinite crystal by a continuum of energy states, or an *energy band*—the π-band.

It is out of place here to give more than a brief account of the theoretical calculations that have been made concerning the band structure of graphite. A detailed survey of the band structure and electronic properties of graphite crystals is given by HAERING and MROZOWSKI[23]. Several papers on the subject appear in the *Proceedings of the 3rd Conference on Carbon*, Buffalo, New York (1959).

Two-Dimensional Model

Because of the relatively weak interaction between the layer planes one possibility is to treat graphite as a two-dimensional solid; that is, to consider only a single layer. The Brillouin zone of graphite contains one electron per atom, and for a single layer the first zone (lower π-band) is a flat hexagon. The second zone (upper π-band) is obtained by extending the sides of the hexagon to form a six pointed star. WALLACE[24] was the first to calculate the band structure of a single layer using the tight-binding method for nearest neighbours. This method starts out from the wave functions of the free electrons, and therefore is analogous to the molecular orbital method applied to solids. Other workers have confirmed WALLACE's general finding that the two π-bands are degenerate in energy at the six zone corners. At these corners, which enclose

the Fermi level, the energy contours are circular and the energy E is a linear function of the wave vector K. Further, in this reduced zone scheme, the E *versus* K curves are symmetrical about the touching energy. *Figure 3* gives a schematic plot

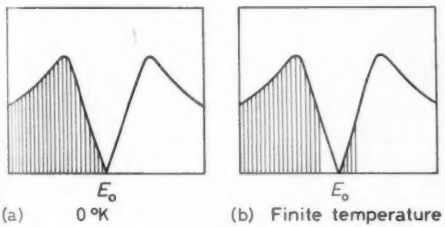

Figure 3. Schematic plot of the density of electron states in the lower (a) *and upper* (b) *π-bands* (*two-dimensional model*)

of the density of electron states $N(E)$ near the band edge (the region of interest is only about 1 per cent of the distance from the zone edge E to the zone centre) at 0°K and at a finite temperature. At 0°K the lower band is completely full while the upper band is empty. At finite temperatures a small fraction of the electrons of highest energy are thermally excited into the upper (conduction) band; this leaves an equal number of positive hole carriers in the lower (valence) band. This general energy scheme places graphite intermediate between a metal and a true semiconductor, but one way of considering graphite is that in the direction of the layer planes it is an intrinsic semiconductor with a zero energy gap[24].

While this simple model gives a good qualitative description of many of the properties of graphite, for example the anomalously high diamagnetic susceptibility is in accordance with the small effective masses of electrons and holes near to band edges, it gives the incorrect prediction that the Hall coefficient of normal graphite should be zero (symmetrical energy contours) and that the electrical conductivity should decrease to zero at 0°K (with one band full and the other empty there are no mobile electrons). In fact, the Hall coefficient is negative under most conditions (predominant electron conduction) and the a-axis conductivity of single crystal graphite increases rapidly at low temperatures.

Three-Dimensional Model

The next stage of the band structure calculations included the interlayer interaction. In this three-dimensional case the Brillouin zone is a flat hexagon bounded by six vertical (100) planes and two horizontal (002) planes. In his 1947 paper WALLACE[24] showed that the main features of the two-dimensional model are retained in the three-dimensional case, but that there were important modifications of the band structure in the region of the six vertical Brillouin zone edges. CARTER and KRUMHANSL[25] investigated the influence of lattice symmetry restrictions (all atoms are not in the same crystallographic position) and found that the density of electron states at the lower edge of the upper band is greater than that of the 'hole' states at the top of the lower band. This could account for the observed negative Hall coefficient of graphite. A refinement by JOHNSON[26] to include higher order effects revealed extra degeneracies near the zone edges, trigonal warping of the energy surfaces and an overlap between the valence and conduction bands. The latter means that there are conduction electrons and holes present even at 0°K; the observed increase in conduction at low temperatures is therefore related to the increased mobility of the current carriers due to decreased thermal scattering. Later SLONCZEWSKI and WEISS[27] abandoned the tight binding approach and performed a group-theoretical, perturbation calculation with considerable success. They concentrated on the influence of lattice symmetry on the electron band structure.

Correlation of Theory with Experiment

The most recent technique in the investigation of the band structure has been the attempts to correlate structure models with various experimental data. These investigations may be divided into two quite distinct approaches to the problem. The first is that of HAERING and WALLACE[28] using the original three-dimensional model of WALLACE[24].

The basis of this model is an unusually low value of the interplanar exchange integral γ_1, which implies that graphite is essentially two dimensional. The value of 0·005 eV for γ_1 was chosen to fit the constant part of the diamagnetic susceptibility. The second striking feature of the model is that ζ_0, the Fermi level at low temperatures measured from the degeneracy corner, is chosen to be $\sim$0·06 eV as derived from the two periods in the de Haas–Van Alphen effect. This value implies that there are excess free electrons ($\sim 10^{-4}$ per carbon atom) present even at zero temperature—a fact which has not yet received a satisfactory explanation. The in-plane exchange integral γ_0 is chosen as $\sim$2·5 eV which is in close agreement with other theoretical and experimental estimates. Using only the three parameters HAERING and WALLACE are able to explain most of the properties of graphite in external fields. Also the model predicts that the ratio of the principal conductivities of graphite is 10^5 in fair agreement with certain experimental values.

The second approach by McCLURE[29] and others is based on the SLONCZEWSKI and WEISS[27] model. It is more involved in that it incorporates six basic energy parameters, and the presence of both holes and electrons as a result of the vertical overlap of the π-bands. The simultaneous presence of holes and electrons is in direct agreement with the findings of cyclotron resonance studies[30]. The six parameters are evaluated from or supported by data obtained from measurements of the magnetic susceptibility and the de Haas–Van Alphen effect, the electronic specific heat, cyclotron resonance, and infra-red absorption studies[31]. Here γ_1 is much larger ($\sim 0{\cdot}14$ eV) than in the HAERING and WALLACE model but the value of γ_0 (~ 3 eV) is similar. Since in the McCLURE model free current carriers are a natural consequence of the overlap of bands it may appear that this model is more satisfactory than the 'small γ_1 model', but in fact both models are successful in explaining a large number of experimental results. Neither model explains *all* the known phenomena, and it may be anticipated that the ideal band structure will incorporate certain features of both existing models.

Doubtless in the near future there will be a considerable advance in our knowledge of the theoretical aspects of the electronic properties of graphite. This should follow when more precise experimental data are available for graphite containing controlled amounts of altervalent impurities, and possibly for rhombohedral graphite[32]. One aspect that especially needs theoretical and experimental attention is the mechanism of electrical transport in the *c*-axis direction of graphite. At present, experimental evidence is hindered by the unknown effect of various types of crystal disorder on *c*-axis properties.

Non-Ideal Graphite

The above description of the electron band structure of graphite refers to the ideal crystal of infinite extent. Most specimens found in practice fall short of the ideal structure due to the presence of defects, as described in *Part I*[1]. The various types of defect may be divided into two groups; those which directly affect the π electron distribution in graphite, and those which primarily cause scattering of the conduction electrons. It is the first group which interests us here.

The change in the π electron distribution, through an increase in the number of electrons in the conduction band (electron donor defects, *e.g.* boron substituted in the lattice[1]) or of holes in the valence band (*e.g.* electron trapping defects[1]) has a marked influence on electronic properties since in ideal graphite these properties are determined by the relatively small concentration of free conduction electrons and holes. Even small defect concentrations of around 1 in 10^4 carbon atoms can cause a significant change both in the number and in the relative predominance of the two types of carrier. This basic scheme is also thought to apply in the case of the electronic properties of dilute crystal compounds of graphite and of irradiated graphite (to be described in *Part III*). It is possible, however, that in the more heavily doped graphites, where there is a pronounced change in the interlayer structure, changes in electronic properties may arise from an actual change in the structure of the Brillouin zones.

Electrical and Magnetic Properties of Graphite

Space does not permit a comprehensive discussion of the various electronic properties of graphite. The aim of this section is merely to outline some of the important properties of the ideal crystal and to indicate briefly how these properties are influenced by crystal imperfections and by the random orientation of crystallites in the case of polycrystalline graphite. Some of the electronic properties not discussed are cyclotron resonance[30], infra-red absorption[31], paramagnetic resonance absorption, the thermionic constants and the dielectric and optical properties.

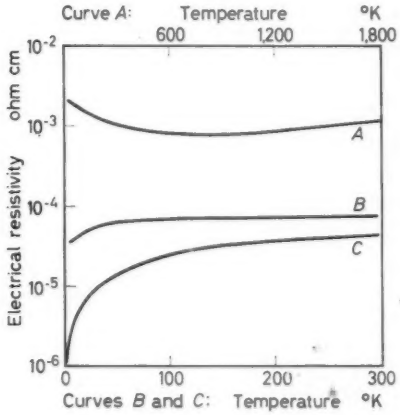

Figure 4. Typical plots of the temperature variation of the electrical resisitivity of polycrystalline graphite (A), pyrolytic graphite (B) and a-axis single crystal graphite (C). Note the different temperature scale of (A)

Electrical Transport Properties

The anisotropy of the electrical resistivity of ideal graphite exceeds that of any other property. Although precise measurements on small single crystals are difficult there is general agreement that the room temperature electrical resistivity in the direction of the layer planes is $\sim 4 \times 10^{-5}$ ohm cm and that the temperature coefficient of resistance is positive to the lowest temperatures. Values reported for the *c*-axis resistivity, however, are not in close agreement and there is some doubt as to the temperature coefficient below room temperature. The resistivity varies from around 1 ohm cm[33,34] to 10^{-2} ohm cm[35] for natural graphite, and is about 0·3 ohm cm for pyrolytic graphite[36]. This variation is not clearly understood at present but it is presumably related to differences in the crystal perfection of the specimens. For example, the presence of spiral dislocations could tend to give too low values by providing relatively low resistance '*a*-axis' paths between layers[1]; conversely, microscopic cracks between layers would greatly increase the the resistance.

The electrical resistivity of natural and synthetic polycrystalline graphite is intermediate between the *a*- and *c*-axis values for a single crystal and usually lies in the range 8 to 12×10^{-4} ohm cm. Such extreme values are typical of block graphite parallel and perpendicular to the axis of extrusion; this bulk anisotropy accords with the fact that conduction in polycrystalline material is predominantly *via a*-axis paths. The temperature coefficient of resistance of polycrystalline graphite is negative from below room temperature up to several hundred degrees centigrade, when it changes to positive. The position of the minimum resistance tends towards lower temperatures as the size or the degree of perfection of the crystallites increases. The negative region has been attributed both to the presence of intercrystallite potential barriers and to the existence of a finite energy gap between the valence and conduction bands of structurally imperfect graphite[1]. *Figure 4* illustrates the variation with temperature of the electrical resistivity of polycrystalline, pyrolytic and single crystal graphites.

A very sensitive test of the approach to ideality of a graphite sample is the transverse magneto-resistance, *i.e.* the relative change of resistance $\Delta\rho/\rho_0$ in the presence of an external transverse magnetic field. For near-perfect single crystals the magnitude of $\Delta\rho/\rho_0$ in the *a*-axis direction is extremely large and increases rapidly with decreasing temperature. For one specimen SOULE[37] reports a change from 5·77 at room temperature to 8250 at 4·2°K in a field of 23,000 oe. Less perfect and polycrystalline specimens show very much smaller values while incompletely graphitized carbons sometimes show a negative magneto-resistance[38]. With such a variation between samples it is perhaps

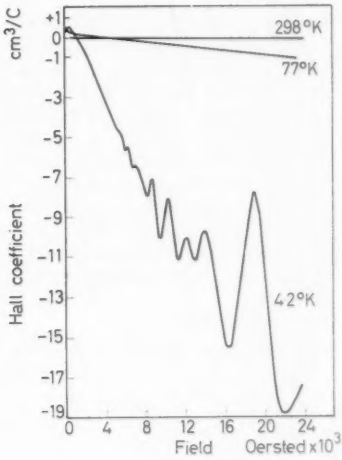

Figure 5. Temperature and magnetic field dependence of the Hall coefficient of graphite (SOULE[37]. *Courtesy Phys. Rev.*)

strange that the field dependence of the magneto-resistance of most graphites obeys the empirical relation.

$$\Delta\rho/\rho_0 = BH^{1\cdot76\pm0\cdot02}$$

where B is a parameter dependent on the temperature and the crystal perfection of the graphite.

The Hall coefficient of graphite is also sensitive to the degree of crystal perfection. For near-ideal crystals the room temperature Hall coefficient is $\sim -0{\cdot}07$ cm³/Coulomb but for defective graphites, in which the Fermi level is displaced into the valence band electron trapping at structural or chemical defects, the value is either much smaller or is positive due to the predominance of hole carriers.

At low temperatures and high fields the Hall coefficient of near-ideal crystals becomes very large, and, as in the case of the magneto-resistance, de Haas–Van Alphen field fluctuations are observed (*Figure 5*). At low fields there is a change in the sign of the Hall coefficient to positive; SOULE[37] sites this as evidence for the presence of fast minority current carriers previously postulated from cyclotron resonance measurements[39].

Thermoelectricity

The thermoelectricity of graphite has recieved comparatively little attention to date (especially theoretically). As mentioned in *Part I* well graphitized carbons have a negative thermoelectric power associated with predominant electron conduction. Recently BLACKMAN, DUNDAS and UBBELOHDE[36] investigated the thermoelectric power of a wide variety of graphites and concluded that not only is this parameter anisotropic, as is the case with the hexagonal metals, but that it differs in sign in the direction of the two principal crystallographic directions. To what extent the positive *c*-axis thermoelectricity is due to positive carriers or to the effect of ' phonon-drag ' or Umklapp processes is not known at present. The anisotropy of thermoelectric power has lead to the construction of an all-graphite thermocouple capable of operating to around 3000°C in a non-oxidizing atmosphere[39].

Diamagnetic Susceptibility

The diamagnetic susceptibility of graphite in the direction perpendicular to the layer planes is anomalously high. Using large single crystals GANGULI and KRISHNAN[40] obtained the values (in c.g.s. units $\times 10^{-6}$)

$$\chi_c = -22{\cdot}0 \qquad \chi_a = -0{\cdot}5$$

and found that interplanar susceptibility was strongly temperature dependent, characteristic of an electron gas with a low degeneracy temperature. The general features of the susceptibility have been explained in terms of the highly anisotropic Brillouin zone structure of graphite, and there have been several theoretical calculations based on both two and three dimensional models. At very low temperatures de Haas–Van Alphen fluctuations are observed having the same field periodicities as found with the magneto-resistance and Hall coefficient. This effect is associated with the majority hole and electron carriers. SOULE[37] has combined the de Haas–Van Alphen data from diamagnetic susceptibility, magneto-resistance and Hall coefficient measurements with that from cyclotron resonance studies and concludes that majority holes and electrons are present in about equal numbers ($\sim 10^{-5}$ per carbon atom) and that the effective masses in the basal plane are $0{\cdot}06m$ for holes and $0{\cdot}03m$ for electrons.

References

[1] BLACKMAN, L. C. F. *Research, Lond.* (1960) **12,** 390
[2] NERNST, W. *Ann. Phys.* (1911) **36,** 395
[3] MAGNUS, A. *Ann. Phys.* (1923) **70,** 303
[4] HOVE, J. E. *Industrial Carbon and Graphite.* London: Soc. Chem. Ind. 1958, 501
[5] DE SORBO, W. and NICHOLS, G. E. *J. Phys. Chem. Solids* (1958) **6,** 352
[6] DE SORBO, W. and TYLER, W. W. *J. chem. Phys.* (1953) **21,** 1660
[7] BOWMAN, J. C. and KRUMHANSL, J. A. *J. Phys. Chem. Solids* (1958) **6,** 367
[8] KRUMHANSL, J. A. and BROOKS, H. *J. chem. Phys.* (1953) **21,** 1663
[9] BERMAN, R. *Proc. phys. Soc., Lond.* (1952) **A65,** 1029
[10] POWELL, R. W. *Industrial Carbon and Graphite.* London: Soc. Chem. Ind. 1958, 46
[11] BOWMAN, J. C., KRUMHANSL, J. A., and MEERS, J. T. *Industrial Carbon and Graphite.* London: Soc. Chem. Ind. 1958, 52
[12] SMITH, A. W. *Phys. Rev.* (1954) **95,** 1095
[13] TYLER, W. W. and WILSON, A. C. *Phys. Rev.* (1953) **89,** 870
[14] HOVE, J. E. and SMITH, A. W. *Phys. Rev.* (1956) **104,** 892
[15] EULER, J. *Naturwissenschaften* (1952) **39,** 568
[16] CURRIE, L. M., HAMISTER, V. C. and MACPHERSON, H. G. *International Conference on Peaceful Uses of Atomic Energy.* New York: United Nations (1956) **8,** 451
[17] NELSON, J. B. and RILEY, D. P. *Proc. phys. Soc., Lond.* (1945) **57,** 477
[18] KELLEY, K. K. *U.S. Bureau of Mines Bulletin No. 383* (1935)
[19] BASSETT, J. *J. phys. Rad.* (1939) **10,** 217
[20] KERN, D. M. *J. chem. Educ.* (1956) **33,** 272
[21] HALL, H. T. *Proc. 3rd Carbon Conference.* Buffalo, New York, 1959, 75
[22] BUNDY, F. P., HALL, H. T., STRONG, H. M. and WENTORF, R. H. *Nature, Lond.* (1955) **176,** 51
[23] HAERING, R. R. and MROZOWSKI, S., to be published in *Progress in Semiconductors* New York: John Wiley & Sons
[24] WALLACE, P. R. *Phys. Rev.* (1947) **71,** 622
[25] CARTER, J. L. and KRUMHANSL, J. A. *J. chem. Phys.* (1953) **21,** 2238
[26] JOHNSON, D. F. *Proc. Roy. Soc.* (1955) **A227,** 349; (1956) **A237,** 48
[27] SLONCZEWSKI, J. C. and WEISS, P. R. *Phys. Rev.* (1958) **109,** 272
[28] HAERING, R. R. and WALLACE, P. R. *J. Phys. Chem. Solids* (1957) **3,** 253
[29] MCCLURE, J. W. *Phys. Rev.* (1957) **108,** 612
[30] GALT, J. K., YAGER, W. A. and MERRITT, F. R. *Proc. 3rd Carbon Conference.* Buffalo, New York. 1959, 193
[31] BOYLE, W. S. and NOZIÈRES, P. *Phys. Rev.* (1958) **111,** 782
[32] HAERING, R. R. *Canad. J. Phys.* (1958) **36,** 352
[33] KRISHNAN, K. S. and GANGULI, N. *Nature, Lond.* (1939) **144,** 667
[34] DUTTA, A. K. *Phys. Rev.* (1953) **90,** 187
[35] PRIMAK, W. and FUCHS, L. H. *Phys. Rev.* (1954) **95,** 24
[36] BLACKMAN, L. C. F., DUNDAS, P. H. and UBBELOHDE, A.R. *Proc. Roy. Soc.* (1960) **A255,** 293
[37] SOLUE, D. E. *Phys. Rev.* (1958) **112,** 698 and 708
[38] MROZOWSKI, S. and CHABERSKI, A. *Phys. Rev.* (1956) **104,** 74
[39] UBBELOHDE, A. R., BLACKMAN, L. C. F. and DUNDAS, P. H. *J. Soc. Chem. Ind.* (1959) 595
[40] GANGULI, N. and KRISHNAN, K. S. *Proc. Roy. Soc.* (1941) **A177,** 168

THE SCOPE OF ORGANOTIN COMPOUNDS IN INDUSTRY

ERNEST S. HEDGES

Director, International Tin Research Council

New fields of application are accessible to tin through recent researches into the synthesis and properties of organic compounds of this metal. Some of these substances are powerful fungicides which are finding use in the protection of crops, wood, textiles, paper and paint against attack by fungi and other pests.

ORGANOMETALLIC compounds, in which a metal atom is directly linked with one or more carbon atoms, have come into particular prominence in recent years, partly because they offer some fascinating new structures to research chemists, but also because some of them have unusual and sometimes unexpected properties.

From the chemist's point of view, organotin compounds have a special interest compared with other organometallic compounds, for tin, like carbon and silicon, takes its place in the fourth group of the periodic table. The conception of an organotin compound as a substance in which a tin atom is directly linked with a carbon atom can therefore be extended to a somewhat grander conception of organotin compounds in which a tin atom *takes the place* of a carbon atom. This view can be justified by the large number of organotin compounds now known and their relation to the compounds of carbon, and by the existence of compounds in which two or more atoms of tin are linked together. From these thoughts the fascinating prospect follows of developing a whole new chemistry based on tin, just as the whole of organic chemistry is based on carbon and the chemistry of silicones on silicon.

The above considerations provide one half of the reasons why in 1949 the International Tin Research Council initiated the systematic study of organotin chemistry, following preliminary exploration in 1937 which was interrupted by the war. The other half was concerned with finding new uses for tin outside the established metallurgical range. The enormous new chemical industries of this century had so far used very little tin and it seemed clear that tin was more likely to be assimilable in chemical industries if presented in a chemical form and especially in an organic chemical form.

The development of the dual theme as envisaged required simultaneously the exercise of theoretical and practical outlooks. The Council was fortunate in enlisting, more than ten years ago, the enthusiastic collaboration of Professor VAN DER KERK, Director of the Institute for Organic Chemistry, T.N.O., Utrecht. His personal contribution to the success of the work is enormous and over the years he has built up in his laboratory at Utrecht a team of research workers, supported by the International Tin Research Council, who work solely on organotin chemistry.

In addition to the main programme of research at Utrecht the Council has also sponsored research on certain aspects of organotin chemistry in the Universities of Leicester, London, Southampton and Toulouse and in other research establishments in several European countries. In recent years interest in organotin chemistry has spread widely and the Council collaborates with universities, official institutes, and the manufacturers and users of organotin compounds in many countries.

The Main Types of Organotin Compounds

Tin can act either as a bivalent or a quadrivalent atom. In the first case it forms a series of dialkyltins or diaryltins. An example is diethyltin, $Sn(C_2H_5)_2$ which was prepared by Löwig as early as 1852 and has the distinction of being the first organotin compound to be described in chemical literature. Nevertheless, the bivalent organotin compounds have found so far no place in industry and they will not be further considered here.

Shortly after Löwig's experiments with diethyltin, Frankland prepared tetraethyltin, $Sn(C_2H_5)_4$, the prototype of the quadrivalent organotin compounds. In these compounds there are clearly four types possible, represented by R_4Sn, R_3SnX, R_2SnX_2 and $RSnX_3$, where R is an organic radical attached to Sn through a carbon atom and X is a radical which is not linked through a carbon atom; X is typically a halogen or an acid radical linked through oxygen. The radical R is not restricted to simple alkyl or aryl hydrocarbon groups, but may contain functional groups. An important object of recent research has been to devise methods of introducing functional groups into organic radicals attached to tin atoms.

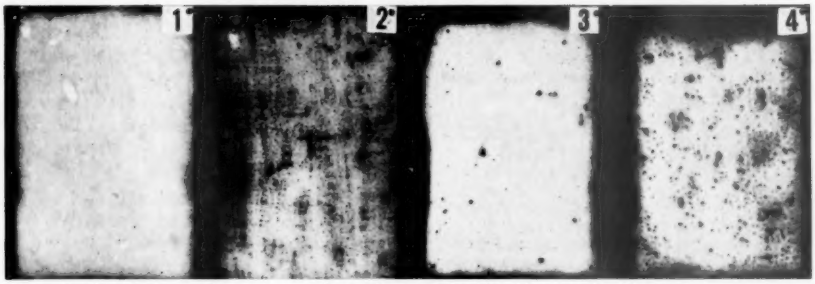

Figure 1. Development of rust on steel panels painted with chlorinated rubber paint and exposed to natural weathering:

1. With dibutyltin dilaurate. Exposure 21 months.
2. Without dibutyltin dilaurate. Exposure 21 months.
3. With dibutyltin dilaurate. Exposure 52 months.
4. Without dibutyltin dilaurate. Exposure 52 months.

The R_4Sn compounds are mostly interesting as starting points for manufacture of the other types. The R_3SnX and the R_2SnX_2 compounds are already commercially used and are referred to below. The $RSnX_3$ compounds, hitherto of little interest, may become more important; some of them are water-repellent and some form the basis of polymers which may prove to be interesting.

Industrial Applications

Dialkyltin Compounds

The greatest tonnage of organotin compounds still goes into polyvinyl chloride (p.v.c.) plastics, an application which was developed in U.S.A. from 1942 onwards. At the temperature reached in the processing of p.v.c. enough decomposition occurs to produce an unsightly brown or black colour in the plastic. This is overcome by adding one or two per cent of a 'stabilizer', a function for which several compounds serve; but when it is necessary for the plastic to preserve a water-white aspect or a fine colour, organotin compounds are used, generally as dibutyltin laurate or maleate or thioglycollate. Other dibutyltin compounds are used and for special purposes the dioctyltin compounds are available.

Much has been written on the chemical mechanism of the stabilizing action, which is certainly complex[1]; but among other things the tin compound prevents the liberation of hydrogen chloride from the plastic during processing and so avoids the autocatalytic decomposition of p.v.c. that would otherwise occur. Tetraphenyltin has a similar effect in preventing the autocatalytic decomposition of synthetic chlorinated transformer oils and is used for that purpose in U.S.A.

Chlorinated rubber is subject to a similar kind of decomposition. Paints formulated with this medium decompose when exposed to sunlight and some hydrochloric acid is liberated, which attacks the structure that has been painted. At the Tin Research Institute it was found that the addition of as little as 0·03 per cent of dibutyltin dilaurate to the paint prevented or retarded this decomposition[2] (*Figure 1*). Chlorinated rubber paints stabilized by organotin compounds are being manufactured in England and have been used on the exterior of some of the largest blocks of flats in London and in swimming pools up and down the country.

An entirely different class of use of dibutyltin dilaurate is as an antihelminthic in the treatment of parasitic worm infections in chickens and turkeys, for which purpose it is used in large quantities in the United States[3]. Further research on the antihelminthic properties of some organotin compounds with respect to mammals as hosts is being carried out in England and in France.

Trialkyl- (or aryl-)tin Compounds

The most outstanding research so far in the field of organotin chemistry is that of VAN DER KERK and his co-workers on the biocidal properties of certain groups of trialkyl- (or aryl-)tin compounds[4]. For example, the lower trialkyltin and triaryltin compounds are found in general to be powerful fungicides. This property of the R_3Sn radical is surprising in view of the non-toxicity of metallic tin. That the action is confined to the tri-substituted compounds is made clear in *Table 1*, where the fungicidal effect of the four types of organotin compounds is compared in relation to their effects on four typical fungi.

The influence of the size of the alkyl groups is, however, an important one, as shown by *Table 2*. Here it will be seen that there is a maximum activity with *n*- or *iso*propyl or butyl substituents in a series of trialkyltin acetates. In general it has come to be recognized that maximum activity is attained when

the total number of carbon atoms in the alkyl groups attached to the tin atom is between nine and twelve. It is, of course, not necessary for all three alkyl radicals to be identical; for example, ten carbon atoms are present in the compounds of dimethyloctyltin. By special methods of synthesis, functional groups can be introduced into the alkyl groups, but hitherto such compounds have not shown any marked superiority over the simpler trialkyltin compounds as regards fungitoxicity. In passing it may be mentioned that many of the organotin compounds are also powerful insecticides.

From the commercial point of view, derivatives of tributyltin and triphenyltin are at present the most important and their production is measured in hundreds of tons.

In the agricultural field, which is the greatest opening for fungicides, considerations of phytotoxicity and of toxic residues make the introduction of new materials a slow process. Nevertheless, triphenyltin acetate has already shown outstanding results in Germany and some other European countries in the control of potato blight and certain fungal diseases of sugar beet and celeriac[5]. With organotin compounds the question of toxic residues is fortunately simplified, for the toxicity resides in the R_3Sn radical, which is shown to be broken down by weathering to R_2Sn, and the eventual material left in the soil is harmless SnO_2. This is in contrast with toxic compounds of copper, lead, arsenic, mercury, *etc.* where the metal itself has a toxic value which persists.

Tributyltin oxide is used in the paper-making industry to combat the bacterial slimes which are apt to accumulate at vital points in the industrial plant. Often they are used in conjunction with other slimicides, such as phenylmercury acetate. Clearly there are opportunities for similar use in many plants involving recirculation of industrial water.

As a bacteriostat, suitably formulated preparations of tributyltin oxide are coming into use in hospitals[6] as floor polishes, washes for the walls and as a stage in the laundering of bed linen, surgical masks, *etc.* In the United States a formulation containing a tributyltin compound is used to prevent the spreading of infections in race-horse stables.

The rot-proofing and moth-proofing of textiles by certain trialkyltin compounds has been demonstrated experimentally (*Figure 2*), but application in this large and diverse industry has been slow, partly because of the degradation to lower alkyltin compounds which occurs in ultra-violet light, with consequent loss of fungitoxicity. The very property

Table 1. Influence of number of alkyl groups bound to tin

	Concentration (mg/l.) which completely arrests development of fungi			
	Botrytis allii	*Penicillium italicum*	*Aspergillus niger*	*Rhizopus nigricans*
Tetraethyltin	50	>1000	100	100
Triethyltin chloride	0·5	2	5	2
Diethyltin dichloride	100	100	500	200
Monoethyltin trichloride	>1000	>1000	>1000	>1000
Stannic chloride	>1000	>1000	>1000	>1000
Stannous chloride	>1000	>1000	>1000	>1000

Table 2. Influence of number of carbon atoms

	Concentration (mg/l.) which completely arrests development of fungi			
	Botrytis allii	*Penicillium italicum*	*Aspergillus niger*	*Rhizopus nigricans*
Trimethyltin	20	20	200	200
Triethyltin	1	2	5	2
Tri-*n*-propyltin	0·1	0·1	1	1
Tri-*iso*-propyltin	0·1	0·1	1	0·5
Tributyltin	0·1	0·1	0·5	0·5
Trihexyltin	1	10	20	100
Trioctyltin	>100	>100	>100	>100
Phenylmercury bromide	0·1	0·1	0·1	2
Phenylmercury acetate	0·5	0·5	0·5	5

which is so useful in agricultural applications is here a disadvantage. Trialkyltin compounds are, however, incorporated in some underclothing, where they have the effect of preventing the bacterial decomposition of sweat, which otherwise leads to the production of bad odours.

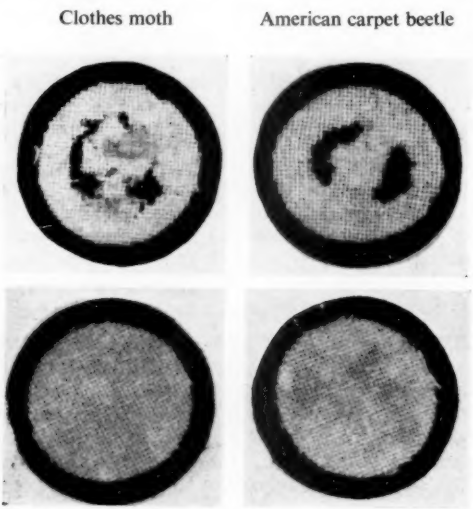

Figure 2. Protection of wool treated with a trialkyltin compound from attack by larvae of clothes moths and American carpet beetles. Above, untreated; below, treated with 0·25 per cent trialkyltin compound

Tributyltin oxide is incorporated in paints, particularly those of the water-emulsion type to prevent the formation of unsightly moulds which are apt to grow on the painted surfaces in damp situations, such as kitchens and bathrooms. Triphenyltin compounds are also incorporated as toxin in anti-fouling paints applied to ships' bottoms.

A clear case for the application of a trialkyltin compound is in the preservation of wood, given the advantages of protection against rot and wood-destroying insects without impairing the ability to take a coat of paint. Early demonstrations of the effectivity of the treatment were made by the Tin Research Institute in the treatment of wooden props in a tin mine in Cornwall and in other situations*. A formulation is now obtainable commercially.

Conclusion

The preceding pages demonstrate how the pursuit of a systematic research, organized by the International Tin Research Council, has in the course of ten years or so with the collaboration of other centres of research and industry, produced a variety of new industrial uses which are steadily growing in importance. Much more may be expected of organotin compounds in the coming years and in this respect special regard may be paid to the probable development of organotin polymers.

References

[1] VERITY SMITH, H. *The Development of the Organotin Stabilizers*, Tin Research Institute, 1959; *British Plastics* (1954) **27**, 176, 213, 307; MACK, G. P., *Kunstoffe* (1953) **43**, 94; WARTMAN, L. H. *Industr. Engng Chem.* (1955) **47**, 1013

[2] GAY, P. J. *Tin and its Uses* (1953) **29**, 6

[3] KERR, K. B. and WALDE, A. W. *Experimental Parisitology* (1956) **5**, 560; KERR, K. B. *Poultry Science* (1952) **31**, 328

[4] VAN DER KERK, G. J. M. *et al. J. appl. Chem.* (1954) **4**, 301, 307, 314; (1956) **6**, 49, 56, 93; (1957) **7**, 356, 366, 369; (1959) **9**, 106, 113, 176, 179

[5] HÄRTEL, K. *Tin and its Uses* (1958) **43**, 9

[6] HUDSON, P. B., SANGER, G. and SPROUL, E. E. *J. Amer. med. Assn* (1959) **169**, 1549

* See *Research* (1960) XIII, 116

RESEARCH ESSAY COMPETITIONS

Mr GREGORY's essay (p. 422) won the first prize in the Waverley Gold Medal Essay Competition. The other prizewinning essays were as follows:

2nd Prize: *The Polybloc System of Construction for Anisotropic Materials* by A. HILLIARD of Société le Carbone-Lorraine.

Special (Under-Thirty) Prize: *Magnetohydrodynamic Power Generation* by R. HAWLEY of King's College, University of Durham.

The winners of the School Essay Competition were as follows:

1st Prize: *The Science of Comfortable Rail Travel* by D. J. EWINS of Kingswood Grammar School, near Bristol.

2nd Prize: *An Investigation into the Effects of Exposure to Wave Action on the Vertical Distribution, and on the Size of Three Species of Fucus* by Miss Shelagh GILL of Calder High School, Liverpool.

SHIPBORNE INERTIAL NAVIGATION SYSTEMS

H. T. HEAL

Royal Naval Scientific Service, Admiralty Compass Observatory, Ditton Park, Slough

The two basic instruments used for inertial navigation are the accelerometer and the gyroscope, and this article describes them and how they are applied to ship-borne inertial navigation systems.

THE basic instruments used in dead reckoning are the log and the compass which give respectively the magnitude and direction of the ship's velocity through the water. Dead reckoning is the art of determining the ship's change of position by integrating this velocity after making corrections for the water current velocity. For this purpose a knowledge of the current velocity as a function of time and position on the earth's surface is required.

The basic instruments used for inertial navigation are the accelerometer and the gyroscope which give respectively the magnitude and direction of an acceleration. The acceleration so measured is not the true rate of change of velocity of the ship as the accelerometer is affected by gravity. It is, however, a simple enough matter to predict accurately the magnitude and direction of the gravitational 'acceleration' as a function of time and position on the earth's surface or, for that matter, anywhere in space. We could thus correct the accelerometer reading at any time, much as the log reading is corrected, and thence by double integration find the change in the position of the ship. This process of integration inevitably introduces errors which increase with time. Inertial Navigation systems of this type are indeed used for missile guidance, where the times involved are measured in minutes, and such systems have been described from time to time in the technical press; they are not, however, well suited to the navigation of ships where the times involved may be measured in hours or even in days.

A ship-borne inertial navigation system operates by determining the orientation of the local vertical at the ship with respect to reference axes fixed in space. This of course is the basis of astro-navigation in which the fixed stars provide the reference axes, and the direction of the vertical is determined either by observation of the horizon or by the use of some gravity-sensitive device such as the 'bubble' of a bubble sextant. Thus, while the missile-borne inertial navigation system may properly be regarded as a form of 'dead-reckoner', the ship-borne system is much more analogous to astronomical navigation.

In navigating a ship on the earth's surface by observing the direction of gravity, two reference directions fixed in space are required. The earth's rotation can be used to provide one of these, namely the direction of the earth's spin axis, but reliance has to be placed on what is in effect a free gyroscope to provide a second reference direction. The net effect of this is that ship-borne inertial navigation systems suffer from a cumulative error in the determination of longitude but provide an indication of latitude which is free from cumulative error.

Accelerometers

Newton's second law of motion states that the rate of change of momentum of a body, that is, its acceleration, is proportional to the applied force. An accelerometer utilizes this law, and by measuring the forces acting on a body of known mass provides a measure of the acceleration of the body. Force and acceleration are of course both vector quantities and may be specified by their components along three mutually perpendicular axes. An accelerometer can be built which measures all three components of acceleration, such an instrument being termed a three-axis accelerometer. Two-axis and single-axis instruments are, however, more common and probably the single-axis accelerometer is the most widely used. A number of accurate accelerometers employ the 'force feed-back' principle, a single-axis accelerometer of this type being constructed as follows. The instrument consists of an outer case, conveniently imagined as cylindrical, inside which is a mass constrained to move along the axis of the case. The constraints are such that they exert no force along the axis. Any movement of the mass relative to the case is detected by an electrical pick-off and the resulting signal is amplified to produce a current which is passed to some form of transducer which exerts force on the mass, tending to restore it to its original position. The transducer may, for example, take the form of a

moving-coil loudspeaker element. The combination of pick-off, amplifier and transducer forms an electrical 'spring' which centres the mass along the case axis and the instrument is designed with this spring so stiff that the mass and case can be taken as moving as one. By measuring the current in the transducer the force acting on the mass along the axis is known and hence the acceleration of the mass, and so of the accelerometer itself, is known. Frequently fluid is introduced to damp the relative motion of the mass and case, or the required damping may be produced by appropriate design of the amplifier.

The chief difficulty in the design of such accelerometers lies in providing restraints to guide the mass, whilst exerting no force upon it along the axis. In the most commonly used suspension system the mass forms the bob of a pendulum which swings either in jewels and pivots or on a weak spring hinge. The mass is thus constrained to move on an arc of a circle, but the permitted motion is so small that the departure from a straight line is negligible. When jewel and pivot suspension is used flotation may be employed, as in the flotation gyroscope described below, in order to reduce the pivot friction.

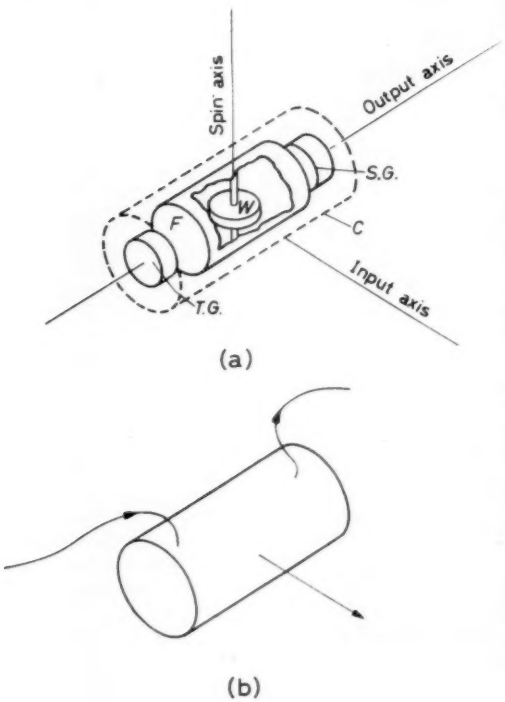

Figure 1.

Gyroscopes

The gyroscope is based on the 'rotational' equivalent of Newton's second law, to wit, 'that the rate of change of angular momentum of a body is equal to the applied torque'. This leads to the well-known effect that the rate of precession of a gyroscope is proportional to the torque applied and that the precession takes place about an axis perpendicular both to the spin axis of the wheel of the gyroscope and the axis about which the torque is applied. As an accelerometer enables the acceleration of a body to be determined by measuring the force acting upon it, so does a gyroscope enable its angular rate with respect to space to be measured by measuring the torques acting upon it. Just as in an accelerometer the major problem lies in suspending the measuring mass so that no unmeasured extraneous forces act on it, so in the gyroscope the main problem lies in ensuring that no unmeasured torques act upon it. Such torques may arise by friction at gimbal bearings, by mass unbalance of gimbal assemblies, by elastic restraint of flexible leads that convey current to the motor that keeps the wheel spinning, and by torque reactions from any electrical pick-offs that are used to detect the orientation of the gyroscope.

The most successful gyroscope design to date is the single-degree-of-freedom flotation gyroscope pioneered by Dr Draper of the Instrumentation Laboratory of the Massachusetts Institute of Technology. This is the so-called Rate Integrating Gyroscope described by W. R. SYMONS* in *Research*. *Figure 1* (*a*) shows diagrammatically such a gyroscope. The wheel *W*, driven by a small hysteresis motor, is mounted within a cylindrical float *F*, with its spin axis normal to that of the float. The float itself is journalled about its axis in an outer concentric cylindrical case *C* and the space between the float and the case is filled with a fluid, the density of the float being adjusted so that it is neutrally buoyant in the fluid. There is thus no load on the journal bearings, either due to the weight of the float or due to any translational acceleration of the case. The bearings are simply pivots and jewels, of the ring-and-endstone variety, and have very little friction. The viscosity of the fluid combined with the use of a very narrow annular gap between the float and the case introduces considerable viscous damping between these two members.

* See 'Miniature Gyroscopes' by W. R. SYMONS, *Research* (1960) **XIII**, 42, the first article in the *Microinstruments* series.

At one end of the case is fitted a torque motor, or torque generator, *T.G.*, which enables known torques to be applied to the float. At the other end is a signal generator, or pick-off, *S.G.*, which

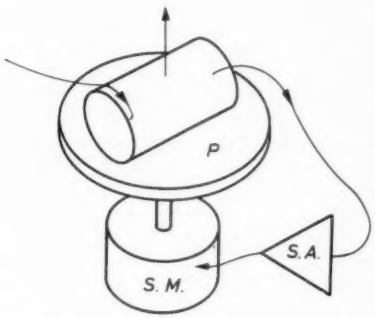

Figure 2.

produces an electrical signal proportional to the rotation of the float with respect to the case. The axis of rotation of the wheel is termed the ' spin ' axis, the float axis is termed the ' output ' axis, and an axis normal to the spin and output axes is called the ' input ' axis. Although the spin and input axes are actually associated with the float and rotate with it, in practice the gyroscope is always used as a null instrument and we can regard the float as always occupying the same position relative to the case, *i.e.* that position in which the signal generator output is zero. The spin and input axes may thus for all practical purposes be associated with the case itself. When it is not required to show the construction of the gyroscope, *Figure 1*(*b*) is a sufficient representation. The ' wires ' at each end of the gyroscope case convey electrical signals into and out of the instrument; that with the arrow into the gyroscope represents the input to the torque generator and the other represents the signal generator output lead. The straight arrow out from the centre of the case shows the direction of the input axis.

A floated gyroscope may be turned into an accelerometer by removing the wheel from the float and inserting an equivalent out-of-balance mass. The float then becomes the pendulum, and the torque and signal generators respectively provide the required transducer and pick-off. It is not uncommon to manufacture gyroscopes and accelerometers which are identical except for the contents of the float.

Figure 2 shows how a floated gyroscope is used to control the movement of a platform. The gyroscope is mounted on a platform *P*, which can be rotated about the input axis of the gyroscope by means of the servo-motor, *S.M.* This motor is supplied with current from a servo-amplifier, *S.A.*, into which is fed the output from the signal generator of the gyroscope. Without going into detailed analysis of this servo-system, its operation may be explained qualitatively as follows. Let a certain torque be applied to the gyroscope by means of its torque generator; then if the gyroscope were free it would precess about its input axis at a certain rate proportional to the applied torque. If by chance the platform is rotating under the action of the motor *S.M.* at this particular rate, then all the torque from the torque generator is used to precess the gyroscope and there is none left over to disturb the *status quo*. If the platform is rotating at a rate less (or greater) than that at which the gyroscope would precess under the applied torque, then there is an excess (or deficit) of torque about the output axis of the gyroscope so that the float rotates slightly with respect to the case until the amplified output from the signal generator has sufficiently speeded up (or retarded) the motor *S.M.* to restore the balance. The net effect is that the platform *P* is power-driven by means of the motor *S.M.*, about the gyroscope input axis, at a rate accurately proportional to the torque applied by the torque generator.

Figure 4 shows, among other things, how a platform may be given known angular rates about three mutually perpendicular axes *X*, *Y* and *Z*, these being the input axes of the three gyroscopes *X*, *Y* and *Z* respectively, all of which are carried on the platform *P*. The arrangement of the *Z* gyroscope, the motor *S.M.Z.* and the amplifier *S.A.Z.* is exactly as in *Figure 2*. It will be noted, however, that the outputs of the signal generators of the *X* and *Y* gyroscopes are connected to the amplifiers '*S.A.* Roll ' and ' *S.A.* Pitch ' through a resolver *R*. This resolver is needed as the *X* and *Y* axes do not necessarily coincide with the axes of the motors ' *S.M.* Roll ' and ' *S.M.* Pitch '. The overall effect is that the platform *P* is power-driven by means of the three servo-motors at a rate the components of which about the *X*, *Y* and *Z* axes are respectively proportional to the torques applied by the torque generators of the corresponding gyroscopes. The component rates of rotation of *P* are of course rates with respect to space. In particular, if no torques are applied by the torque generators then the platform is non-rotating in space, independent of any rotation of the base (deck) on which the system is mounted. We say that the system provides ' three-axis base motion isolation ' of the platform *P*.

The 84-Minute Pendulum

In *Figure 3*, O is the earth's centre and P is a point on the earth's surface so that OP is the vertical through P and the length of OP is R, the earth's radius. Pivotted at P is a simple pendulum with arm of length l and a bob of mass m. Also pivotted at P, and rigidly connected to the simple pendulum, is a wheel of moment of inertia I. The wheel itself is balanced but the combination of wheel and simple pendulum forms a compound pendulum. The moment of inertia of the simple pendulum about P is assumed to be negligible compared with that of the wheel. The angle that the pendulum makes with the vertical is θ. OS is a fixed datum line, where S may, for example, be a star, and the angle ϕ between OP and OS is a measure of the position of P. In particular, $RD^2\phi$ is the tangential acceleration of P, where D denotes differentiation with respect to time. Since R is long compared with l, we may also take $RD^2\phi$ to be the acceleration of the bob. The force that acts on the bob to produce this acceleration gives rise to a torque reaction of magnitude $mlRD^2\phi$ Cos θ acting clockwise on the wheel. The weight of the bob produces an anticlockwise torque of magnitude mlg Sin θ where g is the gravitational 'acceleration'. If θ is small then

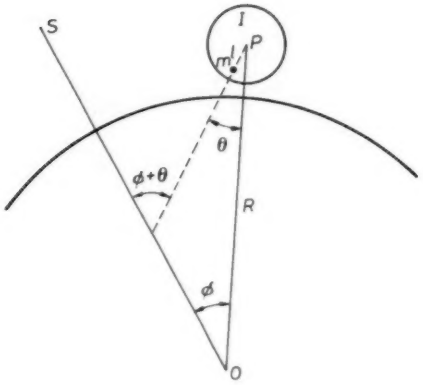

Figure 3.

the net clockwise torque acting on the wheel is $ml\,(RD^2\phi - g\theta)$. The clockwise angular acceleration of the wheel is $D^2(\phi + \theta)$ and the product of this and the moment of inertia of the wheel must equal the applied torque. Hence

$$ID^2(\phi + \theta) = ml(RD^2\phi - g\theta)$$

If the pendulum is proportioned so that $mlR = I$, then the terms in ϕ cancel out and we have left

$$(D^2 + g/R)\theta = 0$$

The behaviour of such a pendulum is thus independent of the tangential acceleration of the point of support, and the pendulum if set swinging will oscillate with a period of $2\pi(R/g)^{\frac{1}{2}}$ which has a value of approximately 84 minutes. If initially vertical, the pendulum will remain so no matter how it is accelerated over the earth's surface. This property of such a pendulum was first stated by Schuler and the pendulum is sometimes called a Schuler pendulum or frequently simply an 84-minute pendulum.

A compound pendulum with such a long period cannot be realized in practice, but it is possible to make a 'synthetic' pendulum having the required properties by the use of gyroscopes and accelerometers. The angle through which a wheel has rotated with respect to space is proportional to the second integral of the applied torque, which in the case considered above was the reaction torque of a simple pendulum. Now the angle through which a gyroscope has precessed is proportional to the first integral of the precessing torque. Therefore if the wheel I is replaced by a gyroscope mounted so that it can precess about the same axis as that about which the wheel rotated, and if the simple pendulum is retained and means are provided for measuring its reaction torque and producing a signal proportional to the integral of this, then, if a torque proportional to this signal is used to precess the gyroscope, the combination will behave as a compound pendulum. In practice torque is not applied to precess a gyroscope by brute force, but a power driven platform is used as described above. A pendulous accelerometer is, of course, precisely an instrument which measures the torque reaction of a simple pendulum, although it is not essential that the accelerometer should be of the pendulum variety.

Figure 4 shows such a synthetic pendulum. The accelerometers AX and AY are sensitive to accelerations along the X and Y axes respectively, that is they detect tilts about the Y and X axes respectively, so that the torque input X_{in} to the X gyroscope is derived through an integrator from the accelerometer AY and the input Y_{in} is derived from AX. It will be noticed that the integrators have a direct path in parallel with them. An undamped pendulum if set swinging will swing for ever and this direct path provides the required damping. Its action is analogous to coupling the simple pendulum and the wheel in the mechanical pendulum of *Figure 3* through a viscous damper instead of locking them solidly together. An integrator shunted by a direct path is not the only 'transfer

function ' that can be used, and other transfer functions may give a somewhat different type of damping. The one here considered is, however, probably the simplest to understand.

Although an undamped 84-minute pendulum is unaffected by lateral acceleration, a damped pendulum (of any type) is affected to a certain extent. The errors in vertical indication so produced are a function of the lateral acceleration and usually an attempt is made to measure the lateral acceleration by some independent means, such as by differentiation of the ship's speed as measured by the log, and feeding in a signal to the system so as to eliminate the error.

The Three-Gimbal System

Let us first note the values of the integrator outputs X_{in} and Y_{in} in *Figure 4* when the synthetic pendulum has settled with the Z axis vertical. Then the accelerometer axes are horizontal so that the inputs to the integrators are zero and their outputs are constant. Also the Z axis is remaining vertical so that the platform P must be rotating with respect to space about the axes X and Y at the same rate as the components of the earth's rotation about these axes. Hence the signals X_{in} and Y_{in} are proportional to the components of the earth's rate about the X and Y gyroscope input axes respectively.

So far nothing has been said about the input Z_{in} to the Z gyroscope. It will be noticed that this signal is derived from the signal X_{in} *via* an integrator shunted by a direct path. If a steady state can exist then it must be such that X_{in} is zero, for otherwise Z_{in} is not steady, but since X_{in} is proportional to the component of the earth's rate about the X axis, it follows that this axis is east–west; for this is the only horizontal direction about which there is no component of earth's rate. Furthermore, if a steady state exists, the X axis must remain east–west and this requires that the platform P rotates about the Z axis at a rate equal to the component of the earth's rotation about the Z axis, that is, the signal Z_{in} which is held at the output of the integrator is proportional to the component of earth's rate about the Z axis. It can be shown that, provided certain restrictions on the values of the time constants of the integrators and the gains of the direct paths are observed, then the system will indeed settle into a steady state with the X, Y and Z axes East, North and vertically up respectively, and that the signals Z_{in} and Y_{in} provide a measure of the components of the earth's rate of rotation about the vertical and the North horizontal respectively.

Such a system as shown in *Figure 4* is usually termed a three-axis stabilizer but with some additional apparatus to perform certain computations it can be made into what is known as a three-gimbal inertial navigation system. The ratio of Z_{in} to Y_{in} is

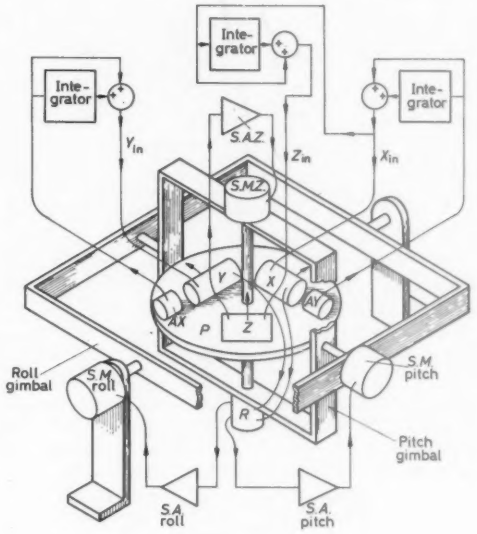

Figure 4.

the ratio of the vertical to the horizontal component of the earth's rate, which is a function of latitude; in fact (Z_{in}/Y_{in}) is the tangent of the latitude. Note that there is no integration involved in the computation of latitude. (There are, of course, three integrators in *Figure 4*. A steady drift of these integrators does not, however, give rise to a steady drift in Y_{in} and Z_{in}, but only produces a constant error.) Also the quantity $(Y_{in}^2 + Z_{in}^2)^{\frac{1}{2}}$ is a measure of the magnitude of the resultant of the horizontal and vertical components of earth's rate which, of course, has a value of one revolution per sidereal day. So far we have considered the system at rest on the earth. Now it can be shown that, at least at moderate latitudes where the ship's speed relative to the earth is small compared with the earth's peripheral speed due to its rotation, the effect of the ship's motion over the earth produces only a small error in latitude indication. Such error as is produced can be largely eliminated by the use of a signal derived from the ship's log. Also, if the ship's longitude is changing, by increasing, say, to the eastward, then the effect on the system is the same as if it were at rest on the earth and the earth were

rotating faster. The increase in the quantity $(Y_{in}^2 + Z_{in}^2)^{\frac{1}{2}}$ above its value with the ship at rest is thus a measure of rate of change of longitude and this must be integrated to give total longitude change.

The Five-Gimbal System

In a five-gimbal system the assembly shown in *Figure 5* replaces the platform *P* of *Figure 4*. The system then has a total of five gimbals, but for clarity the roll and pitch gimbals have been omitted from *Figure 5*. The outermost gimbal shown in *Figure 5*, the azimuth gimbal, carries the two accelerometers and the east-seeking gyroscope which corresponds to the *X* gyroscope of *Figure 4*. Inside the azimuth gimbal is the latitude gimbal which carries the latitude-seeking gyroscope. The input axis of the east-seeking gyroscope is normal to the azimuth gimbal axis and parallel to the latitude gimbal axis, and the input axis of the latitude-seeking gyroscope is normal to that of the east-seeking gyroscope. Finally a longitude gyroscope is gimballed in the latitude gimbal with its gimbal axis along its own input axis and normal to the input axes of the other two gyroscopes.

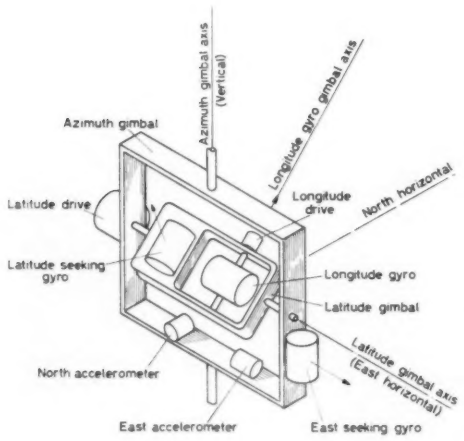

Figure 5.

When the assembly of *Figure 5* is installed in the two outer gimbals of *Figure 4*, then base motion isolation of the assembly can be achieved by passing the outputs of the signal generators of the three gyroscopes through a suitable chain of resolvers and servo-amplifiers to the motors *S.M.* Roll, *S.M.* Pitch and *S.M.Z.* The introduction of the latitude gimbal requires not only a resolver on the azimuth gimbal axis as shown in *Figure 4*, but also an additional resolver on the latitude gimbal axis. This additional resolver is not shown in *Figure 5*.

It will be seen from *Figure 5* that the latitude gimbal is driven round with respect to the azimuth gimbal by means of a 'latitude drive'. Similarly a 'longitude drive' rotates the longitude gyroscope with respect to the latitude gimbal. These two drives are actually two high-gain position servos which force the angles between the gimbals to follow the geared-down output shafts of two velodyne integrators. The high-gain, and hence stiff, servos ensure that torque can be transmitted from gimbal to gimbal without causing any undesired relative rotation. This is essential for successful operation of the base motion isolation system. In spite of their similarity the two drives produce somewhat different effects. Application of a signal to the latitude drive velodyne causes the output shaft to rotate at a proportional rate and the associated position servo causes the latitude gimbal to rotate in sympathy with respect to the azimuth gimbal. Since the east-seeking gyroscope, in conjunction with the remainder of the base motion isolation system, ensures that the azimuth gimbal does not rotate about the latitude gimbal axis, then operation of the latitude drive simply leaves the azimuth gimbal fixed in space whilst the latitude gimbal moves round with respect to it. Similarly, operation of the longitude drive causes the longitude gyroscope to move with respect to the latitude gimbal about the longitude gimbal axis. But the base motion isolation system does not permit rotation of the longitude gyroscope about its input axis. Hence operation of the longitude drive leaves the longitude gyroscope non-rotating in space and instead causes the latitude gimbal to rotate about the longitude gimbal axis, carrying the azimuth gimbal with it.

Application of signals to the torque generators of the east-seeking and latitude-seeking gyroscopes will cause the azimuth gimbal to rotate about the input axes of these two gyroscopes.

If the assembly of *Figure 5* is aligned with the input axis of the east-seeking gyroscope, parallel to that of the *X* gyroscope of *Figure 4*, then the latitude and longitude gyroscope input axes are coplanar with those of the *Y* and *Z* gyroscopes. Since we have means of producing known rates about the latitude and longitude gyroscope input axes, by applying signals to the latitude gyroscope torque generator and the longitude drive velodyne respectively, then by applying signals through a suitable resolver we can cause the assembly of *Figure 5* to

have known rates about the vertical and the North horizontal, these rates being respectively proportional to the two inputs to the resolver. The resolver for this purpose is mounted on the latitude gimbal axis but is not shown in *Figure 5*.

The north accelerometer of *Figure 5* is connected through the usual integrator and direct shunt path to the east-seeking gyroscope, thus forming an 84-minute pendulum which levels the azimuth gimbal about the east–west direction. The signal from the east accelerometer is passed through its integrator and shunt path to the appropriate input of the above mentioned resolver so as to cause rotation of the azimuth gimbal about the north–south direction. Thus an 84-minute pendulum is formed which levels the azimuth gimbal about this direction and so brings the azimuth gimbal axis vertical. Also a signal from the east-seeking gyroscope, corresponding to the signal X_{in} in *Figure 4*, is fed through an integrator and shunt path to the remaining input of the resolver so as to cause the azimuth gimbal to rotate and bring the input axis of the east-seeking gyroscope into the true east–west direction.

Three-axis stabilization of the azimuth gimbal is thus achieved, with the azimuth gimbal axis vertical and the latitude gimbal axis along the east–west horizontal. It will be noted that once the system has settled the latitude-seeking gyroscope is the only one that has any input signal to its torque generator, and this signal is a measure of the component of the earth's rotation about the input axis of the latitude-seeking gyroscope. This signal is applied to the latitude drive, thus rotating the latitude gimbal about its axis until the signal is nulled, which occurs when there is no component of the earth's rotation about the latitude gyroscope input axis, that is, when this axis is normal to the earth's spin axis. Since the east-seeking gyroscope also has its input axis normal to the earth's spin axis it follows that the input axis of the longitude gyroscope is parallel to the spin axis of the earth. As the azimuth axis is vertical, the angle between it and the input axis of the longitude gyroscope is the co-latitude, so that latitude may be read off directly from the angle between the latitude and azimuth gimbals. Also, if the system is at rest on the earth, then the longitude gyroscope will rotate with respect to the latitude gimbal at a rate of one revolution per sidereal day, since the longitude gyroscope does not rotate with respect to space. Indeed, its torque generator is not even connected in this particular system. If the ship is changing her longitude then the rate of rotation will be greater or less according as she steams east or west. By observing the change of angle between the longitude gimbals and subtracting from this the contribution due to the earth's rotation, the change in the ship's longitude can be obtained. The contribution due to the rotation of the earth is, of course, simply obtained by the use of a clock.

Conclusion

The 84-minute pendulum provides a practical method of determining the direction of gravity at sea, and since navigation over long periods of time is required in a ship-borne system the causes of cumulative error must be reduced to a minimum. A ship-borne inertial navigation system will thus usually be of the type in which the ship's position is fixed by observing the direction of gravity relative to co-ordinates fixed in space. One such co-ordinate is established by the earth's spin axis, the other may be established by a free gyroscope. The drift of this gyroscope is then the sole source of cumulative error in the system and gives rise to a drift in the longitude reckoning.

For missile guidance purposes where the missile is not confined to move on the earth's surface, and where the accelerations involved may be large compared with gravity, such systems are inapplicable and recourse must be made to the dead reckoning type of inertial navigation system. In this type both the reference axes are established by free gyroscopes and so are liable to drift, but even more serious are the cumulative errors that result from the double integration of acceleration to give change of position. Since, however, the time of flight of a missile is short such systems are feasible.

Of the ship-borne systems, three- and five-gimbal systems have been described. The three-gimbal system is mechanically simple, but since in this system the ship's position is computed from the electric currents in the gyroscope torque generators, the torque–current relationship of these must be extremely stable and currents must be measured to a high order of accuracy. Furthermore, even after measuring these currents, certain admittedly simple computations have to be performed and if these computations are to be automatic some difficulty arises in producing a computer of the necessary accuracy. An advantage of the three-gimbal system is that the gyroscopes are always the same way up, so that if they are directly affected by gravity, due for example to mass unbalance, then the effect of this is constant and can easily be allowed for.

In the five-gimbal system it is not necessary to measure any currents since latitude and longitude are merely read off as angles between gimbals. For this reason such a system is sometimes termed a 'geometric' system in contrast to the three-gimbal system which is an 'analytic' system. The

five-gimbal system, although mechanically more complex than the three-gimbal system, requires less precision auxiliary equipment. A disadvantage is that the gyroscopes tumble with respect to gravity and errors may be introduced due to mass unbalance of the gyroscopes.

A four-gimbal system is also possible in which the longitude gyroscope of *Figure 5* is mounted direct on the latitude gimbal. This has properties intermediate between the three- and five-gimbal systems, being geometric for latitude reckoning and analytic for longitude reckoning. Longitude rate, with respect to space, is obtained by measuring the torque exerted by the longitude gyroscope torque generator and this torque must be determined with extreme accuracy for the system to be useful. No computation need be performed, however, other than the subtraction of a constant to allow for the earth's rotation so as to give longitude rate with respect to the earth.

In the present state of the art longitude reckoning is the weakest point and the choice of a three-, four- or five-gimbal system is determined solely by consideration of the accuracy obtainable in determining longitude. For a gyroscope that behaves in the same fashion any way up, but which has an unstable torque generator, the five-gimbal system would be called for, whereas for a gyroscope which had a reliable torque generator but which required always to be operated the same way up, then the three-gimbal system would be the obvious choice. Which system will be adopted in the future thus depends largely on which features of the gyroscope are capable of improvement.

Published by permission of the Admiralty.

SURVEY

Atomic Energy Research Establishment, Winfrith

Much of the Research Group of the Atomic Energy Authority has been moved to Winfrith Heath, and is most attractively housed. Construction started in September 1957 and has progressed rapidly; during the past year four new buildings have been built, and the number of reactors is increasing all the time: one zero energy high temperature gas-cooled reactor, ZENITH, went critical in December last year, another reactor, NERO (graphite moderated, zero energy) has been moved down from Harwell and will shortly be followed by DIMPLE, and a fourth zero energy (fast) reactor, ZEBRA, is under construction, as is a 10 KW source reactor, NESTOR. In addition there are exponential and sub-critical assemblies, SCORPIO I, SCORPIO II and HELEN, These are all built for research into the physics of different types of nuclear reactors.

BICEP

The Scorpios are part of the British Industries Collaborative Exponential Programme, a collaborative venture undertaken by the industries building nuclear power stations and the Atomic Energy Authority. They are built for studying reactors containing plutonium as well as uranium in the fuel and can be heated to 400°C. Scorpio I will enable an exponential reactor assembly to be used to predict the critical sizes of power reactors after various periods of operation; Scorpio II is smaller, and enables the sub-critical assembly to be used. It gives information on the details of local variations of neutron distribution within the lattice units. HELEN, a sub-critical assembly will assist in extending the calculation methods used for uranium fuelled liquid moderated systems to plutonium fuelled systems.

The Dragon Project

This international project has been sponsored by the European Nuclear Energy Agency of O.E.E.C.; the research and development is being carried out by an international staff. A high temperature gas-cooled reactor is being constructed (see *Figure 1*) and will be operated on an experimental basis. The fuel will be a mixture of uranium, thorium and graphite, contained in the graphite moderator—no metal cans will be used. Thus it is necessary to use an impermeable graphite to avoid excessive activity in the coolant, which is to be helium, and various graphites are being tested for their behaviour under irradiation. The temperature at the centre of the core could be as high as 1500°C, and research is in hand on the materials, to be used, instrumentation, control, heat transfer, mass transfer etc. DRAGON is expected to be operating at full power towards the end of 1963.

The reactor physics of the system is being investigated in the reactor ZENITH, whose core can be heated electrically to produce the conditions anticipated in DRAGON, and it will be used for two years to check reactor physics design calculations for DRAGON.

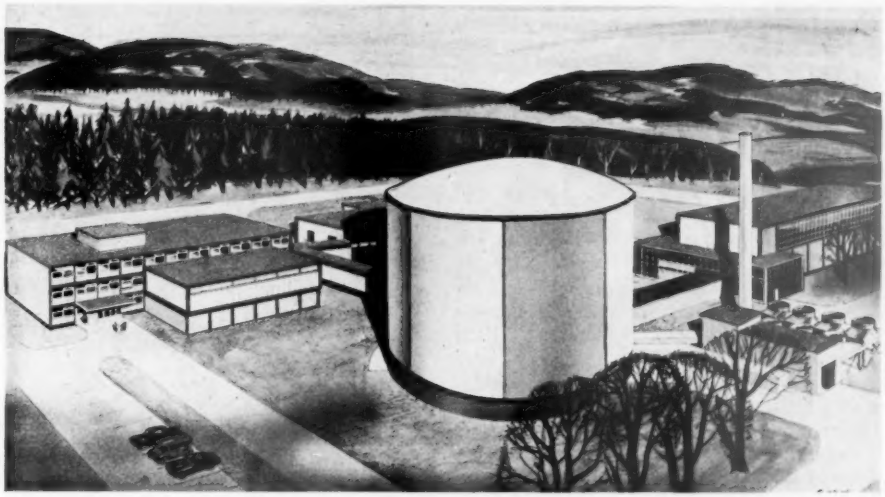

Figure 1. Artist's impression of the completed Dragon Project at Winfrith. (Courtesy U.K.A.E.A.)

Pure and Applied Chemistry

The first issue of the official journal of the International Union of Pure and Applied Chemistry has just been published. It is edited by Professor B. C. L. WEEDON of London University, the Chairman of the Editorial Advisory Board is Dr H. W. THOMPSON, and it is published by Butterworths at £6 per volume. The first issue of *Pure and Applied Chemistry* includes the *Proceedings of the Radioactivity Analysis Symposium* which took place in Vienna in June 1959.

Post Office Research Station

An object of much research at Dollis Hills Post Office Research Station is cable economy, *i.e.* the use of the minimum amount of cable for the maximum number of messages. Various approaches to this problem are being made, several of which were demonstrated at Dollis Hills during their open days:

Synthetic Speech—An analysis–synthesis telephone link was demonstrated in which two kinds of spectral analysis are carried out simultaneously on the speech. The resulting signals, constituting a running description of the short-term speech spectrum, occupy a total transmission bandwidth of less than 200 c/s, compared with the 3100 c/s in ordinary telephone transmission. These signals are then used at the receiving end of the link to control apparatus which produces an adequate, though not yet identical, synthetic copy of the input speech.

Pulse Code Modulation—The multiplexing of speech transmission on existing cable by use of 8-digit pulse code modulation was demonstrated; not only can very many more conversations be carried on one cable using this form of transmission, but it is claimed that the systems are relatively free from effects of noise and cross talk.

Long-distance Circular Waveguides—Looking further into the future, work is in progress on long distance 'piped radio', using a copper pipe 2 in. in diameter and perfectly circular, which will carry at least 100 television channels or 100,000 telephone circuits (*Figure 2*), transmitted as electromagnetic waves confined within the tube. The main difficulty at present is that not only must the pipe be perfectly circular, but it must also be perfectly straight so that no energy is lost. Different methods of connecting pipes and of bending the waves without losing power are being investigated, as it is obviously impractical to expect to be able to lay miles of absolutely straight pipe.

Electronic Letter Sorter

The postal services of the G.P.O. are gradually becoming mechanized. At the open days three years ago an apparatus for facing letters automatically was exhibited and this year we were shown an automatic sorter. The letters are presented to a keyboard operator who taps a code representing the postal district to which they should be routed, thus

printing correctly coded dots of phosphorescent and, to the naked eye, almost invisible material on the envelope. The letters are then fed into a sorting machine which scans them with ultra violet light, thus reading the dot pattern by means of a photocell, and routes the letter to its correct pigeon hole in the machine. This machine is in operation at Luton and will shortly be installed at Mount Pleasant and at Norwich, and it is hoped that during the next two years most of the larger sorting offices in Great Britain will have both facing and sorting machines.

The next stage is to complete the sorting automatically at the destination post office, and the Dollis Hills staff is now working on this. Their plan is that each house shall have a code—a series of letters or numbers which will represent first their postman's district and then their house, rather like a telephone number. Letters addressed to them should bear this code, and then the process of sorting would be similar to the one described above; a keyboard operator would tap a phosphorescent code onto the envelope, and the letters would be scanned and sorted. It is thought that although it might take some time for private houses to adopt this scheme, it should be comparatively easy to persuade businesses to use it if they believed that they would receive their post more rapidly.

Figure 2. Millimetric waveguide laboratory at the Post Office Research Station; measurements are being made here to determine the effectiveness of a component used to connect together tubes of different diameters

Erratum

In Magnetic Properties of Epitaxially Grown Films by O. S. S. HEAVENS (Research (1960) **13,** 404) reference 17 should have read *J. de Phys.* (1960) **21,** 48S, not (1959) **20,** 485.

BOOK REVIEWS

Principles of Optics

M. BORN and E. WOLF
(*xxvi + 803 pp; 10 in. by 6¼ in.*)
Oxford: Pergamon Press. 120*s*

ALTHOUGH this monumental work owes much to Born's well-known textbook 'Optik', it is in no way a translation, but a new and up to date book. It amply fulfils the claim of its subtitle to give an account of the 'electromagnetic theory of propagation, interference and diffraction of light'; in addition to the early methods of HUYGENS, FRESNEL and KIRCHOFF, and the electromagnetic treatments of diffraction by SOMMERFELD and MIE (for a conducting half plane and a conducting sphere respectively), there are accounts of the relatively recent work on partially coherent light, phase contrast microscopy, diffraction theory of aberrations and Gabor's method of imaging by reconstructed wavefronts.

But to keep the volume within bounds it has been found necessary to limit the subjects covered to the spirit of the title '*Principles* of Optics'. Some topics normally fully treated in textbooks of optics thus receive only very cursory mention. For example, determinations of the velocity of light are covered in a single page; also the theory of the concave grating and of the Rowland circle.

A number of chapters have been partly or wholly contributed by other authors—Dr W. I. WILCOCK (a very readable chapter of 114 pages on interference and interferometers), Dr P. C. CLEMMOW (rigorous diffraction theory), Professor A. M. TAYLOR (optics of metals) Dr A. R. STOKES (optics of crystals), Professor A. B. BHATIA (diffraction of light by ultrasonic waves) and Dr P. A. WAYMAN (a short chapter of only eighteen pages on image forming instruments). This varying authorship leads to a certain amount of discontinuity of treatment. But it has the advantage that some topics are discussed from varying aspects sometimes giving, for example, both a mathematical and a physical approach.

This book is by no means a text book for the undergraduate, although the treatment is in general so clear and precise that he may find it profitable to scan some parts, in particular, the historical introduction which summarizes the development of optical ideas from the time of Descartes and Galileo to the present day. But the serious student of optics will find it a great boon to have in one volume so much material for which he has hitherto had to consult several texts and numberless original papers.

The production is admirable; in particular the diagrams are clear and—a point often not conceded to the reader's convenience—the half tone plates are to be found where they are discussed in the text even when this leaves large areas of glossy pages blank. G.F.L.

Encyclopaedia on Cathode Ray Oscilloscopes and their Uses

J. F. RIDER and S. D. USLAN
(*xix + 132 pp; 11½ in. by 9 in.*)
New York: John F. Rider; London: Chapman & Hall. 210*s*

THE reviewer of this book is an ordinary user of oscilloscopes, not an expert on them and it is from the point of view of such a user that the following remarks are made.

First it must be confessed that the book is a large one. When shown to each of seven colleagues, all interested in 'the oscilloscope art' (see front of dust cover), they laughed out loud at the size of the volume. Scientific books grow even bigger and better. This one has a weight of seven and a half pounds (compare a bound Volume 27 of *Review of Scientific Instruments* at 6 lbs) and measures 11½ in. × 9 in. × 3 in. thick. In an age when oscilloscopes are becoming truly portable the encyclopaedias describing them will soon need the trolleys thus rendered obsolete.

The contents of the book are encyclopaedic in scope though there are omissions. For example, there is not a word about the early history of cathode ray tubes nor of their competitors. The text is really intended as a handbook for modern practising workers who wish to apply what they read to the work in hand, rather than a storehouse of knowledge on oscillographs. On this purely technical side there is a vast amount of information, some of which is almost certainly useful to the average oscilloscope user (oscilloscoper?) whatever his field.

After discussing the functioning principles of tubes, including excursions into very elementary electrostatics and relativistic mass change (with an inadequate square root sign in the relevant equation), focusing and screens, the first part of the text carries a good account of the associated circuitry, amplifiers, time bases and the rest. Sections on applications form the middle portion and these are interesting for their useful graphs on square wave testing of R–C coupled networks. The most fascinating part is the long chapter on engineering, medical and scientific applications. There are brief details of the use of oscilloscopes in dentistry, razor blade manufacture, bridge building, coal mining and many other fields. One's own speciality is almost certainly mentioned. Regrettably, and this goes for most of the text, there are in general no references so that further details of bright ideas in one branch cannot easily be obtained by workers in another. The final part is the one the present reviewer would most willingly do without. It contains (according to the front dust cover) complete specifications and schematic diagrams of all American made oscilloscopes. This must be very interesting to manufacturers and anyone running an oscilloscope servicing business, but for the average user, even if his own model is 1959 or earlier, this material must be superfluous unless he has lost the data manual which generally comes with the machine. Maybe in future a free copy of the encyclopaedia (with appendix on the new model) will come instead.

The general verdict is that the book must be of especial interest to practical men in the electronics field, but all users would enjoy a browse. The range in expertize of potential readers for which the encyclopaedia tries to cater is so wide that it is difficult to say how it would strike an interested dentist, T.V. service mechanic or professional scientist. F.C.F.

International Conference on Co-ordination Chemistry (London, April 6th–11th, 1959)

(*204 pp; 8¾ in. by 5½ in.*)
London: The Chemical Society. Special Publication No. 13. 37*s* 10*d*

IN APRIL 1959 the Chemical Society held an international conference on co-ordination chemistry, under the sponsorship of the International Union of Pure and Applied Chemistry. The present volume contains the full text of seven interesting lectures delivered on that occasion and abstracts of well over one hundred papers of more variable interest which were submitted to the conference. The lectures were contributed by K. ZIEGLER ('New Aspects of Some Organometallic Complex Compounds'), H. M. N. H. IRVING ('The Stability of Metal Complexes'), H. W. STERNBERG and I. WENDER ('Metal Carbonyls and Related Compounds as Catalytic Intermediates in Organic Syntheses'), H. TAUBE ('Mechanisms of Complex Ion Reactions: Recent Advances'), E. O. FISCHER ('Co-ordination Compounds of Unsaturated Hydrocarbons with Metals'), L. E. ORGEL ('Metal-Ligand Bonds'), and R. NAST ('Complex Acetylides of Transition Metals'). They occupy over half the space of the book.

This volume is obviously essential reading for anyone engaged in research involving metal compounds, but it will also be of some interest to a wider group of chemists by giving, within a small space, a fair picture of current ideas and activities of inorganic chemists. It is interesting to note that the great economic importance of organometallic compounds as catalysts used in the chemical industry appears to be responsible for much of the impetus to preparative work in this field.

Only a few years ago scientific investigation in this branch of chemistry was largely dominated by X-ray crystallography. Such studies still play an essential role to-day, but lively interest is now also taken in the solution chemistry of co-ordination compounds. Absorption spectra and stability constants of complexes and the kinetics of their reactions in solution (especially oxidation–reduction processes and ligand displacements) are being studied to an increasing extent and with due awareness of the consequences of interionic forces. The range of phenomena to which theoretical ideas on structure and reactivity can be applied has consequently greatly widened and, in turn, has renewed interest in theoretical developments (for example, in crystal field theory and its progeny 'ligand field' theory).

It has rightly been said that recent years have seen a 'renaissance' of inorganic chemistry. The present volume bears witness to the sturdiness of the infant

phœnix who—as healthy babes generally do—likes to give a lusty shout when hungry, in need of love and attention, or wishing to display the vigour of his lungs. V.G.

The Neurochemistry of Nucleotides and Amino Acids

R. O. BRADY and D. B. TOWER (*Eds*)

(*xii + 292 pp; 9¼ in. by 6 in.*)

New York: John Wiley; London: Chapman & Hall. 80*s*.

THE importance of the nervous system to human activities and understanding needs no emphasis today. Neurochemistry is among the most active of the sciences contributing to neurology, and forms a section of the American Academy of Neurology under the Chairmanship of Dr D. B. TOWER who is one of the editors of the present book. The book consists of papers and discussions which were contributed to a Symposium of the Academy, and which are of an admirably high standard and well documented. The majority of the papers are reviews of the contributors' own field of work; although they are grouped round the subjects of amino acids and nucleotides, these compounds make metabolic contact with many other aspects of neurochemical study.

Thus the nucleotides are both structural constituents, as described by D. F. BRADLEY and M. K. WOLF, and intermediary metabolites and coenzymes, as recounted by E. G. TRAMS, D. R. SANADI, R. J. ROSSITER, B. W. AGRANOFF, R. M. BURTON and N. O. KAPLAN. In these roles they contribute to the synthesis of lipid and polysaccharide constituents of the brain, and mechanisms for many of these syntheses are well documented. Amino acids also contribute both intermediary metabolites and structural materials to the nervous system, as recounted by H. WAELSCH with respect to proteins. Aspects of the metabolism of aromatic amino acids are described by S. UDENFRIEND; of asparagine and glutamine by D. B. TOWER, and of γ-aminobutyric acid by C. F. BAXTER, E. ROBERTS and R. W. ALBERS, while the importance of serine in contributing to phospholipids is indicated by D. M. GREENBERG.

In general, the authors are strongest in discussing metabolic aspects of their subject, and to this they make notable contributions. Several comments on electrophysiological phenomena are uncritical, and biochemical aspects other than those of intermediary metabolism are often lacking; this reflects to a considerable extent the position in current neurochemical research as a whole. The Editors are to be congratulated on the excellent production of the book, and on its most valuable index. Too often indexes to symposia are lacking; here it is welcome and gives lasting value to a collection of papers whose wide ramifications will lead them to be consulted by workers in other biochemical and neurological specialities. H. McI.

The Future of Non-Ferrous Mining in Great Britain and Ireland

(*xxvi + 614 pp; 8¾ in. by 9½ in.*)

London: Institution of Mining and Metallurgy. 60*s*

THIS book contains twenty four papers delivered at a symposium held in September, 1958 and a full report of the discussions on each of the papers. In addition there is a useful summary of prospects and recommendations for the future by Mr Frank HIGHAM, a valuable index, and the text of an address on the future of mineral exploration by Dr Anton GRAY.

The ground covered is very comprehensive. Besides geological descriptions of all the important ore-fields there are historical details and assessments of the future potential of each area. There is also a section on exploration techniques and another on economics and finance. As all the authors are the best available experts, both geologists and mining engineers, and because several industrialists concerned in the various mining enterprises contributed to the discussions, this is a highly authoritative book.

The great fall off in mining in the British Isles is due to the virtual exhaustion of the known economic ore-bodies, which are, as might be expected in a country whose geology is known in such detail, all those bodies which can be discovered at the surface. There are, however, within known fields certain bodies which do not reach the surface, and which can only be discovered by cross-cutting. More important, there is good reason to believe that there are certain areas in which there might be many valuable ore-bodies which are quite unknown because they are concealed. Examples are an area in Derbyshire where lead–zinc ores might be expected in the carboniferous limestone which is concealed by younger rocks, another area in Derbyshire where bodies may be concealed by carboniferous lavas, and the area covered by the sea off St Agnes and Cligga in Cornwall where tin ores may be expected. These prospects are discussed in detail in papers by Dr K. F. G. HOSKING and Professor K. C. DUNHAM.

The great problem facing the industry is that it cannot finance the exploration of these prospects without Government assistance and it cannot obtain that assistance until the exploration has been done so that the Government can be convinced that the industry has a future. As regards tin-mining it seems certain that there will be a revival some day; for as Professor W. R. JONES points out, the alluvial and eluvial deposits of the world will sooner or later be worked out and then Cornwall and Devon, one of the few areas of the world in which important assemblages of tin-lodes are known to occur, will come into their own again. S.S.

NOTICE TO CONTRIBUTORS: All communications intended for the Editorial Department should be addressed to The Editor, *Research*, 88 Kingsway, London, W.C.2. Contributions sent on approval will be carefully considered by the Editor; but no responsibility whatsoever can be accepted for manuscript, although every effort will be made to return them if a stamped addressed envelope is enclosed for that purpose. The copyright of all contributions for which a fee is paid shall belong to Butterworth & Co. (Publishers) Limited, together with the right of republication. The Editor reserves the usual right to make alterations in the text of articles accepted for publication. The author will receive galley proofs and if alterations have been made every effort will be made to obtain the author's prior agreement. Unless special arrangements are made, contributions are only received and considered on these conditions.